

# Exhibit M



## The Fundamental Relationship Between Sample Mass and Sampling Variance in Real Geological Samples and Corresponding Statistical Models

C.R. STANLEY<sup>1</sup>

(Received July 14, 2006; accepted April 6, 2007)

**Abstract** — In mineral exploration and mining applications, it is sometimes necessary to reduce sampling error in order to improve representivity of individual sample assays. Historic efforts to reduce sampling error have relied on the belief that larger samples exhibit less sampling error. Justifications for this belief traditionally have coupled the binomial theorem with an ideal geological material consisting of ore and gangue particles of equal size and shape (an equant grain model) and empirical tests of geological materials, to illustrate an inverse relationship between sample size and sampling variance. To date, no proof exists of this sample size–sampling variance relationship in real geological materials exhibiting variations in grain size, shape, composition, and degree of liberation.

Using first principles calculus and the formula for the mean, sample mass and sampling variance are proven to be inversely proportional in all geological materials. Distribution assumptions and physically ideal geological models are not used in this proof. Furthermore, algebraic manipulations of expressions describing the variances of equant grain models consistent with binomial, hypergeometric, and Poisson distributions reveal that these models also exhibit an inverse sample size–sampling variance relationship. Thus, the sampling behaviors of these models are numerically consistent with the sampling behaviors of real geological materials, and these models can be used to estimate sampling errors in real geological materials using simple sampling parameters. © 2007 Canadian Institute of Mining, Metallurgy and Petroleum. All rights reserved.

**Key Words:** Sampling error, sample mass, binomial distribution, hypergeometric distribution, Poisson distribution, equant grain model.

**Sommaire** — En exploration minérale et en production minière, il est quelquefois nécessaire de réduire l'erreur d'échantillonnage afin d'améliorer la représentativité des analyses d'échantillons individuels. Les tentatives historiques de réduction de l'erreur d'échantillonnage reposaient sur la croyance que des échantillons de plus grande taille possèdent une erreur d'échantillonnage moindre. Les justifications traditionnelles de cette croyance découlaient de l'application du théorème binomial à un matériel géologique idéalisé constitué de particules de minerai et de gangue de taille et de forme similaires (un modèle équi-granulaire), et d'essais empiriques sur des matériaux géologiques, pour illustrer une relation inverse entre la taille de l'échantillon et la variance de l'échantillonnage. Il n'existe à ce jour aucune preuve que cette relation entre la taille de l'échantillon et la variance de l'échantillonnage est valide dans des matériaux géologiques réels dont les constituants présentent des variations granulométriques, géométriques, compositionnelles et de degré de libération.

Il est possible de démontrer à partir d'algèbre élémentaire et de l'équation de la moyenne que la masse de l'échantillon et la variance de l'échantillonnage sont inversement proportionnels dans tous les matériaux géologiques. Cette preuve ne fait pas appel aux hypothèses quand à la distribution ou à un modèle géologique physique idéal. De plus, la manipulation algébrique d'expressions décrivant la variance de modèles équi-granulaires possédant des distributions binomiales, hypergéométriques, et Poisson montre aussi une relation inverse entre la taille de l'échantillon et la variance de l'échantillonnage. Ainsi, le comportement à l'échantillonnage de ces modèles est mathématiquement similaire au comportement à l'échantillonnage de vrais matériaux géologiques, et ces modèles peuvent être utilisés pour estimer les erreurs d'échantillonnage dans de vrais matériaux géologiques par le biais de paramètres d'échantillonnage simples. © 2007 Canadian Institute of Mining, Metallurgy and Petroleum. All rights reserved.



<sup>1</sup> Department of Earth and Environmental Science, Acadia University, Wolfville, Nova Scotia, B4P 2R6; e-mail: cliff.stanley@acadiau.ca



## Introduction

Recent work by Smee and Stanley (2005) and Stanley and Smee (2005, in press) demonstrates that sampling error typically represents the largest source of component error in geochemical determinations for Au. In many cases, this sampling error represents greater than 90% of the variance in duplicate samples, and this is largely due to the extraordinarily 'nuggety' character of gold-bearing material. In contrast, component errors introduced during subsampling after crushing, during further subsampling after pulverization, and during geochemical analysis for Au collectively are much lower, and do not typically exceed 25% of the total variance in duplicate samples.

Geochemical determinations for other metals (e.g., Cu, Pb, Zn, Ni, Fe, Al, etc.) typically do not exhibit such large sampling errors, because the ore grains in samples containing these elements are more abundant than in samples containing Au. As a result, sampling error comprises a smaller proportion of total measurement error in base metal-bearing geological materials. However, if proper sample preparation procedures and analytical methods designed to limit the introduction of large sample processing errors are employed, the magnitude of error introduced during these latter stages of geochemical analysis will be small. As a result, even for geochemical determinations of elements not subject to large nugget effects, sampling error likely represents the majority of the error in geochemical analyses. Consequently, reducing this largest component (sampling) error in any geological material is the most efficient means of reducing the total error in the geochemical analyses (Bettenay and Stanley, 2001; Smee and Stanley, 2005; Stanley and Smee, 2005, in press).

Determination of sampling error can be undertaken using 'sampling duplicates' collected in the field, either from outcrop or drill core/cuttings, along with duplicates collected after sample crushing ('crushed duplicates'). The sampling duplicates report an error that is the sum of all of the component errors introduced during each stage of sampling, preparation, and analysis ( $\sigma_1$ ). In contrast, the postcrushing duplicates only report the errors introduced during subsequent preparation and analysis ( $\sigma_2$ ). As a result, subtraction of the error variance described by the postcrushing duplicates from the error variance described by the sampling duplicates yields an estimate of the error variance introduced during the initial collection of the geological material (the sampling error:  $\sigma_s^2 = \sigma_1^2 - \sigma_2^2$ ; Shaw, 1961; Kleeman, 1967; Royle, 1985; Smee and Stanley, 2005; Stanley and Smee, 2005, in press).

Reducing sampling error has traditionally been achieved by the collection of larger samples (Wilson, 1964; Kleeman, 1967; Ingamells et al., 1972; Gy, 1982; Pitard, 1989a,b, and references therein regarding Gy's sampling theory). This strategy is based on the belief that sampling variance is inversely proportional to sample mass. This belief was originally justified using binomial statistics applied to the sampling of an ideal, synthetic model of a geological material (an equant grain model, or EGM; Herden, 1960; Wilson, 1964; Kleeman, 1967; Clifton et al., 1969;

Ingamells, 1981; Stanley, 1998). This equant grain model is assumed to have sampling characteristics similar to the geological material to be sampled. However, it is constrained to be particulate in nature, and consists of particles that are all the same size and shape that are either ore or gangue grains, each of constant composition. Furthermore, the ore grains in the EGM are assumed to occur homogeneously distributed throughout the material to be sampled. Thus, the EGM exhibits an overall grade uniformity, and local grade variations beyond those occurring due to random processes do not occur (i.e., there is no segregation or distributional heterogeneity; Visman, 1969; Duncan, 1971; Visman et al., 1971; Ingamells and Switzer, 1973; Gy, 1982; Miller, 1991; Lyman, 1998; Pitard, 1989a,b, and references regarding Gy's sampling theory, therein). If the above assumptions hold, the resulting sampling error observed in the EGM is described by a random (binomial) process associated with the chance inclusion or exclusion of ore (and gangue) grains in geochemical samples.

Obviously, real geological materials are not all particulate in nature, do not consist of grains of the same size, shape, and composition, are not all fully liberated (consisting of only ore or gangue material), and may not behave during sampling like a binomial variable. As a result, it is not generally known whether sampling variance is inversely proportional to sample mass in all real geological materials. To date, geologists only know that the ideal binomial EGMs thought to simulate the sampling characteristics of real geological materials exhibit this relationship.

It is therefore worthwhile to determine what the relationship is between sampling variance and sample size in real geological materials so that geologists can know, or at least predict, exactly how much sampling variance reduction will result from a specific increase in sample mass. In the following section, the numerical relationship between sampling variance and sample mass in real geological materials is derived in a general way using first principles calculus and simple descriptive statistics. This derivation makes no unrealistic or unverifiable assumptions that prevent general application of the results to all geological materials.

## Sample Size and Sampling Variance

To determine how the mass of a sample of a geological material relates to the sampling variance for an element concentration in the sample, and thus relates to the sampling variance of any geological material under consideration, consider a set of element concentrations ( $c_i$ ) made on  $k$  samples of some small but equal mass ( $m_s$ ) from a geological material. Assume that this geological material exhibits a homogeneous distribution of the element of interest, at least at and above the scale of the sample. If the geological material is particulate in nature, and has been or can be mixed to remove any existing segregations, this assumption is neither unrealistic nor limiting. Note also that provided homogeneity exists, the sample need not be particulate in nature.

Because the geological material exhibits uniformity, the expected values of the concentrations,  $E(c_i)$ , and sampling



variances,  $E(s_{c_i}^2)$ , in each of these small samples are the same ( $E(c_i) = E(\bar{c}_{mr})$ ;  $E(s_{c_i}^2) = E(s_{mr}^2)$ ), where the subscript *mr* refers to these small samples. Thus, given that the mean element concentration is:

$$\bar{c}_{mr} = \frac{1}{k} \sum_{i=1}^k c_i, \quad (1)$$

the formula describing the estimation error on this mean element concentration can be derived by propagating the errors in each  $c_i$  through Equation 1 into  $\bar{c}_{mr}$ . This is achieved using a generative formula derived from a Taylor first order expansion about the mean:

$$\sigma_{\bar{c}_{mr}}^2 \approx \sum_{i=1}^k \left( \frac{\partial f(x_i)}{\partial x_i} \right)^2 \sigma_{c_i}^2 \quad (2)$$

assuming all concentrations are independent (Meyer, 1975; Stanley, 1990). All of the partial derivatives of Equation 1 ( $\partial \bar{c}_{mr} / \partial c_i$ ) required in Equation 2 equal  $1/k$ , and because all  $\sigma_{c_i}^2 = \sigma_{mr}^2$ , the variance on the observed mean becomes:

$$\begin{aligned} \sigma_{\bar{c}_{mr}}^2 &\approx \sum_{i=1}^k \left( \frac{\partial \bar{c}_{mr}}{\partial c_i} \right)^2 \sigma_{c_i}^2 = \sigma_{mr}^2 \sum_{i=1}^k \left( \frac{1}{k} \right)^2 \\ &= \sigma_{mr}^2 \left( \frac{k}{k^2} \right) = \frac{\sigma_{mr}^2}{k}. \end{aligned} \quad (3)$$

Equation 3 is the conventional formula for the square of the standard error on the mean (the standard variance on the mean; Spiegel, 1975). Note that the approximation in Equation 3 becomes an equality when  $\sigma_{mr}^2$  is sufficiently small.

Finally, multiplying both numerator and denominator on the right side of Equation 3 by  $m_r$ , the mass of these small samples, yields:

$$\sigma_{\bar{c}_{mr}}^2 = \frac{\sigma_{mr}^2 m_r}{k m_r} = \frac{\sigma_{mr}^2 m_r}{M_r}, \quad (4)$$

where  $k m_r$  is the combined mass ( $M_r$ ) of the  $k$  small samples taken together.

Now let us consider a single, larger sample with mass  $M$ , as above. The expected value of the concentration of this single large sample [ $E(c_{Mr}) = \mu_{Mr}$ ] is equal to the expected value of the individual concentrations and mean concentration of the original set of  $k$  small samples [ $E(\bar{c}_{mr}) = E(c_i) = \mu_{mr}$ ]. This is because these two parameters ( $\mu_{Mr}$  and  $\mu_{mr}$ ) are derived from large and small samples of the same homogeneous material. Analogously, the expected value of the sampling variance on the observed element concentration of the large sample concentration [ $E(s_{Mr}^2) = \sigma_{Mr}^2$ ] will be equal to the expected value of the standard variance on the mean [ $E(s_{\bar{c}_{mr}}^2) = \sigma_{\bar{c}_{mr}}^2$ ] for the original set of  $k$  small samples, again, because these variances are derived from materials that are mechanically equivalent. Given these large and small sample mean and variance equalities, Equation 4 (which describes the col-

lective sample mass-sampling variance relationship for the original  $k$  small samples) can be converted into Equation 5 (which describes the analogous relationship for the large sample):

$$\sigma_{Mr}^2 M_r = \sigma_{mr}^2 m_r \text{ or } \frac{\sigma_{Mr}^2}{M_r} = \frac{\sigma_{mr}^2}{m_r}. \quad (5)$$

Equation 5 indicates theoretically that an inverse relationship exists between sample mass and sampling variance in all homogeneous geological materials. As a result, Equation 5 can be used to determine the sampling variance ( $\sigma_{Mr}^2$ ) that will occur in samples of some mass ( $M_r$ ) of a geological material, given the sampling variance ( $\sigma_{mr}^2$ ) observed in samples of a different mass ( $m_r$ ) of that same geological material.

Furthermore, the logarithm of Equation 5 is:

$$\log \sigma_{Mr}^2 = -\log M_r + \log (\sigma_{mr}^2 m_r). \quad (6)$$

Thus, on a logarithmically-scaled diagram of sample mass [ $\log M_r$ ] versus sampling variance [ $\log \sigma_{Mr}^2$ ], Equation 6 describes a straight line with an intercept of  $\log (\sigma_{mr}^2 m_r)$ , and a slope of -1 (Figure 1). The intercept of this line,  $\log (\sigma_{mr}^2 m_r)$ , represents a single fundamental sampling parameter for any geological material under consideration, precisely because it alone establishes the relationship between sample mass and sampling variance for a real geological material. Thus, by plotting the observed sampling variance against the mass of samples used to estimate that variance on such a diagram, one can deduce the value of this fundamental sampling parameter ( $\log (\sigma_{mr}^2 m_r) = \psi_r$ ; which has units of mass; cf. the  $K_s$  of Ingamells and Switzer, 1973) by determining the logarithm of the sampling variance of this line at a unit sample mass (in whatever mass units apply; Figure 1).

Note that the above derivation of the sample mass-sampling variance relationship does not rely on assumptions about the distribution of the sampling errors, nor does it require any simplifying assumptions about the physical characteristics of the sample (e.g., particulate character, size distribution, mineralogy, density, liberation, or shape) other than that it is homogeneous at the scale of the sample size. Furthermore, this derivation is independent of the concentration of the element of interest. Consequently, this inverse relationship between sample mass and sampling variance applies to all sampling errors at all concentrations in all geological materials.

### Equant Grain Model

Historical efforts to estimate sampling error have not generally employed the inverse relationship between sample mass and sampling variance for real geological materials described in Equation 5. Rather, these efforts have typically, and unnecessarily, relied on simplifying assumptions that define an EGM that mimics the sampling char-



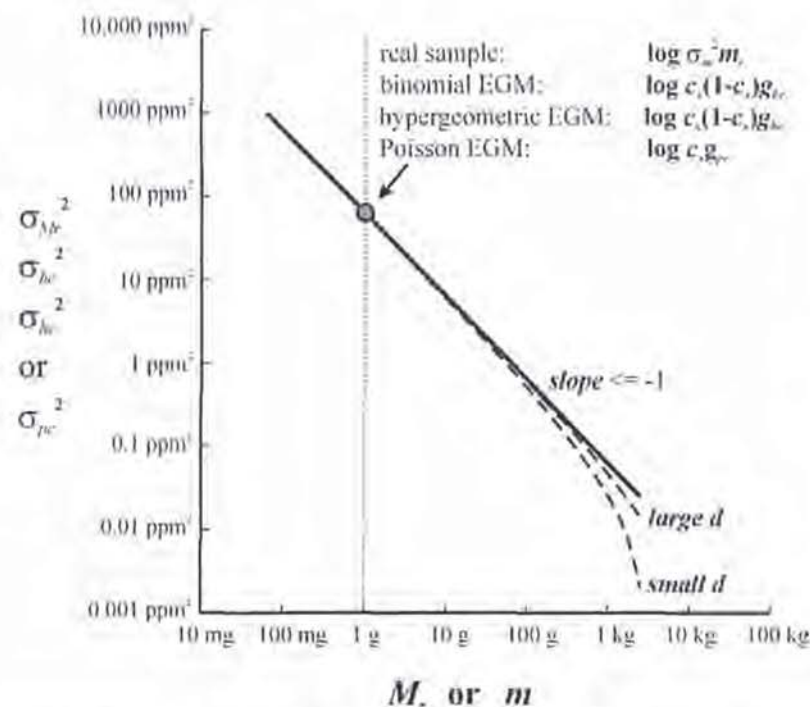


Fig. 1. Logarithmically scaled scatterplot of sample mass ( $m$  or  $M$ ) versus sampling variance ( $\sigma_{br}^2$ ,  $\sigma_{hr}^2$ ,  $\sigma_{pr}^2$ , or  $\sigma_{pc}^2$ ) for a real sample of any geological material exhibiting any element concentration (real sample mass— $M$ , vs. real sampling variance— $\sigma_m^2$ ), and of binomial (sample mass— $m$  vs. binomial sampling variance— $\sigma_{br}^2$ ), hypergeometric (sample mass— $m$  vs. hypergeometric sampling variance— $\sigma_{hr}^2$ ) and Poisson (sample mass— $m$  vs. Poisson sampling variance— $\sigma_{pr}^2$ ) equant grain models. The solid line describes the limiting sample mass-sampling variance relationship and has a slope of -1. The dashed lines describe the sample mass-sampling variance relationship for hypergeometric EGMs with different  $d (= N/n$ ; population size/sample size) parameters. The intercept for the real sample line ( $\log \sigma_m^2$ ) is a function of the small real sample mass ( $m$ ) and the small real sampling variance ( $\sigma_{br}^2$ ), the product of which is itself a real sampling parameter. The intercepts for the binomial ( $\log c_i(1-c_i)g_{br}$ ), hypergeometric ( $\log c_i(1-c_i)g_{hr}$ ), and Poisson ( $\log c_i g_{pr}$ ) EGM lines are functions of the concentration of ore grains ( $c_i$ ) and the equant grain masses for the binomial, hypergeometric and Poisson EGMs ( $g_{br}$ ,  $g_{hr}$ , and  $g_{pr}$ ). These equant grain masses are ideal sampling parameters for each model. All of these sampling parameters (real and ideal) are independent of concentration.

acteristics of the geological material under consideration, to allow application of simple statistical theory to predict sampling error (Herdén, 1960; Clifton et al., 1969; Engels and Ingamells, 1970; Ingamells et al., 1972; Ingamells and Switzer, 1973; Ingamells, 1974a,b, 1981, 1986; Ingamells and Engels, 1976; Cheng, 1995; Sketchley, 1997, 1998; Stanley, 1998, 2003). Fortunately, the sample mass-sampling variance relationship in real samples and EGMs based on the binomial distribution are the same (an inverse relationship).

However, the binomial distribution is not the only distribution employed to estimate sampling error. In fact, a number of different approaches involving two other statistical distributions, the Poisson distribution and hypergeometric distribution, have also been used to estimate sampling error (Clifton et al., 1969; Engels and Ingamells, 1970; Ingamells et al., 1972; Ingamells and Engels, 1976; Ingamells, 1981; Stanley, 1998, 2003). These alternative approaches have similarly and necessarily required analogous simplifying

assumptions (i.e., EGMs) so that the properties of these statistical distributions can be employed to estimate sampling error. Conveniently, these alternative approaches employ the same general set of assumptions (EGMs) to allow application of statistical theory to the problem of modeling sampling errors. These assumptions (EGMs) are summarized below.

An EGM is a fictitious, ideal, synthetic geological material meant to exhibit the sampling characteristics of a real geological sample (Herdén, 1960; Clifton et al., 1969; Ingamells, 1981; Stanley, 1998), but which has physical characteristics that are consistent with the assumptions required by the statistical distribution used to model the sampling error observed in the real geological material. Consequently, an EGM consists of mineral grains that are all of the same mass and shape. These grains are assumed to be of two, fully liberated, types: ore and gangue. To simplify matters, ore grains are assumed to consist solely of the element or commodity of interest to be analyzed (cf., Ingamells, 1974a,b, 1981, 1986). Thus, although a real Cu ore, say from a sandstone-hosted stratabound Cu deposit, may contain both chalcocite ( $\text{Cu}_2\text{S}$ ; 34.63% Cu) and bornite ( $\text{Cu}_5\text{FeS}_4$ ; 63.31% Cu) ore minerals, as well as a certain small amount of non-recoverable Cu in the gangue minerals, the ore grains in the associated EGM are assumed to consist solely of native copper grains (i.e., 100% Cu). Similarly, gangue grains consist only of the

other elements in the geological material that are not of interest, and are assumed to have a constant composition. Because only one element is considered to be of interest at a time, the composition of the gangue grains is not constrained to be equal to the bulk composition of the other elements in the geological material, and thus can exhibit any composition (i.e., they can be any mineral with any density), provided that they contain none of the element of interest (i.e., 0% Cu). It is commonly convenient to assume that the gangue grains consist of a simple mineral that occurs most abundantly in the geological material under consideration (e.g., quartz in the stratabound Cu deposit example) so that a gangue mineral density can be assigned.

The EGM is constrained to exhibit the same element concentration as the real geological material it is meant to represent. Hence, the concentration (or proportion) of ore grains ( $c_i$ ) is constrained such that:

$$c = c_o c_o + (1 - c_o) c_g, \quad (7)$$



where  $c$  is the concentration of the element of interest in the sample,  $c_o$  is the concentration of the element of interest in the equant ore grains (100%), and  $c_g$  is the concentration of the element of interest in the equant gangue grains (0%). Under these assumptions, Equation 7 reduces to:

$$c = c_o \quad (8)$$

Obviously, because the equant ore and gangue grains have different compositions, their densities ( $\rho_o$  and  $\rho_g$ , respectively) will differ, so the bulk density of the EGM ( $\rho_e$ ) will be equal to:

$$\rho_e = c_o \rho_o + (1 - c_o) \rho_g \quad (9)$$

Note that because the equant ore and gangue grains consist of ideal minerals that are different from those within the real geological material under consideration, the bulk density of the equant grain model ( $\rho_e$ ) will not necessarily equal the density of the real geological material ( $\rho$ ). Fortunately, this discrepancy has no impact on the ability of the EGM to simulate the sampling characteristics of real geological materials, provided that only the mass of a geological sample, and not its volume, is considered. In addition, if density considerations are important, then the density of the gangue grains can be defined such that the EGM density (at the concentration of interest) equals the density of the real geological material ( $\rho$ ); it is meant to represent:

$$\rho_g = \frac{\rho - c_o \rho_o}{(1 - c_o)} \quad (10)$$

Furthermore, because the ore and gangue grains have the same mass but have different (assumed) compositions, and thus densities, ore and gangue grains will have different volumes, and thus different dimensions. Assuming these equant grains are spheres, they will have different diameters. However, because mineral densities typically differ by less than an order of magnitude (e.g., in the sandstone-hosted stratabound Cu deposit example, above, the equant native copper ore grains have a density of 8.90 g/cm<sup>3</sup>, whereas the gangue quartz grains have a density of 2.65 g/cm<sup>3</sup>; thus, these densities differ by a factor of 3.68), the volumes of equant ore and gangue grains will also differ by significantly less than an order of magnitude. Furthermore, because the diameter of a spherical equant grain is proportional to the cube root of its volume, the difference between ore and gangue grain diameters will generally be much less than the cube root of 10 (= 2.15; in the example above, the ore and gangue grain diameters will differ by a factor of 1.50). Consequently, although the physical size of equant ore and gangue grains differ in an EGM, the difference is typically small relative to the range of particle sizes observed in real geological materials, which commonly exhibit grain size ranges over several orders of magnitude (e.g., Stanley, 2006a).

Lastly, the size of an EGM sample is typically described by the total number of equant ore and gangue grains in the sample ( $n$ ). Thus, the number of ore ( $x$ ) and gangue ( $n - x$ ) grains equal:

$$x = nc_o \text{ or } n - x = n(1 - c_o) \quad (11)$$

Because the mass of a sample of an EGM is a function of  $n$  ( $m = ng_e = n v_g \rho_g$ ; where  $g_e$  is the mass of the equant grains and  $v_g$  is the average volume of the equant grains), the mass of an EGM sample can be adjusted to ensure that it exhibits the same sampling characteristics as a real geological material by changing the number of equant grains ( $n$ ) in the EGM sample. In this way, the EGM sample mass can be calibrated such that its sampling error ( $\sigma$ ), or relative sampling error ( $\sigma / \mu = CV$ ), is the same as that observed in a real geological material under consideration (Herden, 1960; Clifton et al., 1969; Stanley, 1998).

Examples of three EGMs associated with different statistical distributions (binomial, hypergeometric, and Poisson) are presented in Figure 2. In each, the large square defines the physical limits of the EGM population, whereas the small square defines the physical limits of an EGM sample. These differ according to the assumptions associated with the statistical distribution employed. Similarly, the proportions (concentrations) of equant ore (gray) and gangue (white) grains vary in these three models, also according to the assumptions associated with these statistical distributions.

## Sampling Error Models

Historical efforts to model sampling error have considered a variety of statistical distributions applied to ideal samples from an EGM (Herden, 1960; Clifton et al., 1969; Ingamells, 1981; Stanley, 1998). In the following discussion, the sample mass-sampling variance relationship is investigated for each of three different statistical models commonly used to describe sampling error in geological materials: the binomial, hypergeometric, and Poisson models. In each case, the relationship between sampling variance and sample mass is derived from the original equations for the variance of binomial, hypergeometric, and Poisson variables. These equations, unlike those for the normal distribution, are functions of either the mean concentration of the element of interest, or the mean number of ore grains in an EGM sample. In all cases, an inverse relationship exists (identical to that illustrated by the above proof for real geological materials, and previously illustrated for a binomial EGM; Wilson, 1964; Kleeman, 1967; Ingamells et al., 1972). This indicates that these theoretical models are consistent with the fundamental relationship between sample mass and sampling variance for real geological materials.

### Binomial Distribution

Consider a sample of size  $n$  collected from an EGM population of size  $N$ . The expected value of the concentration of the element of interest in this sample  $E(c)$  will be ( $= X / N$ ), where  $X$  is the number of ore grains in the population. Because only ore and gangue grains exist and either contain or do not contain the element of interest, the probability of collecting a single ore grain at random from the EGM equals this concentration ( $c_o$ ). However, because a sample from the EGM consists of many grains, some ore and some gangue, many grains must be collected from the



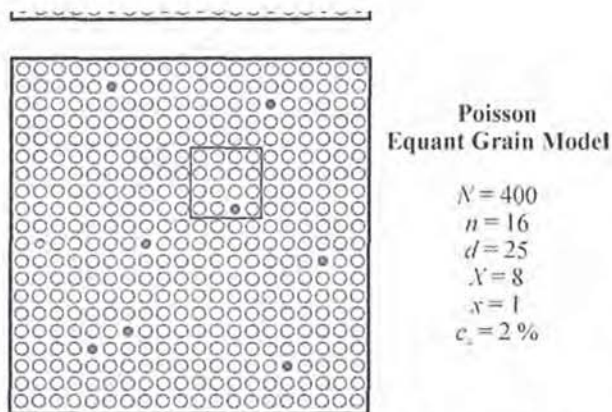


Fig. 2. Schematic examples of various EGMs illustrating the different sampling characteristics of each statistical distribution. Large and small boxes represent the EGM populations and samples, respectively; gray and white circles represent equant ore and gangue grains, respectively.  $N$  = population size;  $n$  = sample size;  $X$  = number of ore grains in the population;  $x$  = number of ore grains in the sample;  $c_s$  = proportion (concentration) of ore grains in the population and the expected number of ore grains in the sample;  $d$  ( $= N / n$ ) can be thought of as the inverse of sampling density.

EGM to obtain a sample. The grains in this sample could be acquired from the EGM by randomly collecting one grain at a time. Although, in reality, geologists would not typically collect the grains in a sample from an EGM in this manner (they would probably scoop out many equant grains at a time from a pile of grains of a particulate material or collect a single or multiple pieces of rock (mineral aggregate) from a non-disaggregated material), statistical considerations make it is worthwhile to consider the sample collected in this way. At each equant grain collection

that represent major proportions of a sample (i.e., they typically comprise greater than say 0.01% of the sample mass; Wilson, 1964; Kleeman, 1967; Spiegel, 1975). The sample mass-sampling variance relationship for a binomial variable can be derived using the equation describing the magnitude of sampling variance for a binomial variable:

$$\sigma_{bx}^2 = nc_x(1 - c_x), \quad (12)$$

(Spiegel, 1975) where  $c_x$  is the concentration (proportion) of equant ore grains in the population,  $b$  refers to the binomial distribution, and  $n$  is the number of equant grains contained within a sample. This equation describes the magnitude of sampling variance in terms of the total number of equant grains in a sample and the concentration of the element of interest contained in the ore grains within the population.

Unfortunately, Equation 12 describes the sampling variance of the number of ore grains within a sample, and does not describe the sampling variance of the concentration of ore grains in a sample. To obtain an equation that describes the sampling variance for the concentration of ore grains, we must rely on the fact that the concentration (or proportion) of ore grains in the sample approximates the concentration of ore grains in the population ( $c_s$ ), and equals the number of ore grains in a sample ( $x$ ) divided by the total number of grains in the sample ( $n$ ):

$$c_s = \frac{x}{n}. \quad (13)$$

Analogous relationships exist between both the mean and standard deviation of the number of ore grains, and the mean and standard deviation of the concentration of ore



## The Fundamental Relationship Between Sample Mass and Sampling Variance in Real Geological Samples and Corresponding Statistical Models

C.R. STANLEY<sup>1</sup>

(Received July 14, 2006; accepted April 6, 2007)

**Abstract** — In mineral exploration and mining applications, it is sometimes necessary to reduce sampling error in order to improve representivity of individual sample assays. Historic efforts to reduce sampling error have relied on the belief that larger samples exhibit less sampling error. Justifications for this belief traditionally have coupled the binomial theorem with an ideal geological material consisting of ore and gangue particles of equal size and shape (an equant grain model) and empirical tests of geological materials, to illustrate an inverse relationship between sample size and sampling variance. To date, no proof exists of this sample size–sampling variance relationship in real geological materials exhibiting variations in grain size, shape, composition, and degree of liberation.

Using first principles calculus and the formula for the mean, sample mass and sampling variance are proven to be inversely proportional in all geological materials. Distribution assumptions and physically ideal geological models are not used in this proof. Furthermore, algebraic manipulations of expressions describing the variances of equant grain models consistent with binomial, hypergeometric, and Poisson distributions reveal that these models also exhibit an inverse sample size–sampling variance relationship. Thus, the sampling behaviors of these models are numerically consistent with the sampling behaviors of real geological materials, and these models can be used to estimate sampling errors in real geological materials using simple sampling parameters. © 2007 Canadian Institute of Mining, Metallurgy and Petroleum. All rights reserved.

**Key Words:** Sampling error, sample mass, binomial distribution, hypergeometric distribution, Poisson distribution, equant grain model.

**Sommaire** — En exploration minérale et en production minière, il est quelquefois nécessaire de réduire l'erreur d'échantillonnage afin d'améliorer la représentativité des analyses d'échantillons individuels. Les tentatives historiques de réduction de l'erreur d'échantillonnage reposaient sur la croyance que des échantillons de plus grande taille possèdent une erreur d'échantillonnage moindre. Les justifications traditionnelles de cette croyance découlaient de l'application du théorème binomial à un matériel géologique idéalisé constitué de particules de minerai et de gangue de taille et de forme similaires (un modèle équigranulaire), et d'essais empiriques sur des matériaux géologiques, pour illustrer une relation inverse entre la taille de l'échantillon et la variance de l'échantillonnage. Il n'existe à ce jour aucune preuve que cette relation entre la taille de l'échantillon et la variance de l'échantillonnage est valide dans des matériaux géologiques réels dont les constituants présentent des variations granulométriques, géométriques, compositionnelles et de degré de libération.

Il est possible de démontrer à partir d'algèbre élémentaire et de l'équation de la moyenne que la masse de l'échantillon et la variance de l'échantillonnage sont inversement proportionnels dans tous les matériaux géologiques. Cette preuve ne fait pas appel aux hypothèses quand à la distribution ou à un modèle géologique physique idéal. De plus, la manipulation algébrique d'expressions décrivant la variance de modèles équigranulaires possédant des distributions binomiales, hypergéométriques, et Poisson montre aussi une relation inverse entre la taille de l'échantillon et la variance de l'échantillonnage. Ainsi, le comportement à l'échantillonnage de ces modèles est mathématiquement similaire au comportement à l'échantillonnage de vrais matériaux géologiques, et ces modèles peuvent être utilisés pour estimer les erreurs d'échantillonnage dans de vrais matériaux géologiques par le biais de paramètres d'échantillonnage simples. © 2007 Canadian Institute of Mining, Metallurgy and Petroleum. All rights reserved.

<sup>1</sup> Department of Earth and Environmental Science, Acadia University, Wolfville, Nova Scotia, B4P 2R6; e-mail: cliff.stanley@acadiau.ca



## Introduction

Recent work by Smee and Stanley (2005) and Stanley and Smee (2005, in press) demonstrates that sampling error typically represents the largest source of component error in geochemical determinations for Au. In many cases, this sampling error represents greater than 90% of the variance in duplicate samples, and this is largely due to the extraordinarily 'nuggety' character of gold-bearing material. In contrast, component errors introduced during subsampling after crushing, during further subsampling after pulverization, and during geochemical analysis for Au collectively are much lower, and do not typically exceed 25% of the total variance in duplicate samples.

Geochemical determinations for other metals (e.g., Cu, Pb, Zn, Ni, Fe, Al, etc.) typically do not exhibit such large sampling errors, because the ore grains in samples containing these elements are more abundant than in samples containing Au. As a result, sampling error comprises a smaller proportion of total measurement error in base metal-bearing geological materials. However, if proper sample preparation procedures and analytical methods designed to limit the introduction of large sample processing errors are employed, the magnitude of error introduced during these latter stages of geochemical analysis will be small. As a result, even for geochemical determinations of elements not subject to large nugget effects, sampling error likely represents the majority of the error in geochemical analyses. Consequently, reducing this largest component (sampling) error in any geological material is the most efficient means of reducing the total error in the geochemical analyses (Bettenay and Stanley, 2001; Smee and Stanley, 2005; Stanley and Smee, 2005, in press).

Determination of sampling error can be undertaken using 'sampling duplicates' collected in the field, either from outcrop or drill core/cuttings, along with duplicates collected after sample crushing ('crushed duplicates'). The sampling duplicates report an error that is the sum of all of the component errors introduced during each stage of sampling, preparation, and analysis ( $\sigma_1$ ). In contrast, the postcrushing duplicates only report the errors introduced during subsequent preparation and analysis ( $\sigma_2$ ). As a result, subtraction of the error variance described by the postcrushing duplicates from the error variance described by the sampling duplicates yields an estimate of the error variance introduced during the initial collection of the geological material (the sampling error:  $\sigma_s^2 = \sigma_1^2 - \sigma_2^2$ ; Shaw, 1961; Kleeman, 1967; Royle, 1985; Smee and Stanley, 2005; Stanley and Smee, 2005, in press).

Reducing sampling error has traditionally been achieved by the collection of larger samples (Wilson, 1964; Kleeman, 1967; Ingamells et al., 1972; Gy, 1982; Pitard, 1989a,b, and references therein regarding Gy's sampling theory). This strategy is based on the belief that sampling variance is inversely proportional to sample mass. This belief was originally justified using binomial statistics applied to the sampling of an ideal, synthetic model of a geological material (an equant grain model, or EGM; Herden, 1960; Wilson, 1964; Kleeman, 1967; Clifton et al., 1969;

Ingamells, 1981; Stanley, 1998). This equant grain model is assumed to have sampling characteristics similar to the geological material to be sampled. However, it is constrained to be particulate in nature, and consists of particles that are all the same size and shape that are either ore or gangue grains, each of constant composition. Furthermore, the ore grains in the EGM are assumed to occur homogeneously distributed throughout the material to be sampled. Thus, the EGM exhibits an overall grade uniformity, and local grade variations beyond those occurring due to random processes do not occur (i.e., there is no segregation or distributional heterogeneity; Visman, 1969; Duncan, 1971; Visman et al., 1971; Ingamells and Switzer, 1973; Gy, 1982; Miller, 1991; Lyman, 1998; Pitard, 1989a,b, and references regarding Gy's sampling theory, therein). If the above assumptions hold, the resulting sampling error observed in the EGM is described by a random (binomial) process associated with the chance inclusion or exclusion of ore (and gangue) grains in geochemical samples.

Obviously, real geological materials are not all particulate in nature, do not consist of grains of the same size, shape, and composition, are not all fully liberated (consisting of only ore or gangue material), and may not behave during sampling like a binomial variable. As a result, it is not generally known whether sampling variance is inversely proportional to sample mass in all real geological materials. To date, geologists only know that the ideal binomial EGMs thought to simulate the sampling characteristics of real geological materials exhibit this relationship.

It is therefore worthwhile to determine what the relationship is between sampling variance and sample size in real geological materials so that geologists can know, or at least predict, exactly how much sampling variance reduction will result from a specific increase in sample mass. In the following section, the numerical relationship between sampling variance and sample mass in real geological materials is derived in a general way using first principles calculus and simple descriptive statistics. This derivation makes no unrealistic or unverifiable assumptions that prevent general application of the results to all geological materials.

## Sample Size and Sampling Variance

To determine how the mass of a sample of a geological material relates to the sampling variance for an element concentration in the sample, and thus relates to the sampling variance of any geological material under consideration, consider a set of element concentrations ( $c_i$ ) made on  $k$  samples of some small but equal mass ( $m_i$ ) from a geological material. Assume that this geological material exhibits a homogeneous distribution of the element of interest, at least at and above the scale of the sample. If the geological material is particulate in nature, and has been or can be mixed to remove any existing segregations, this assumption is neither unrealistic nor limiting. Note also that provided homogeneity exists, the sample need not be particulate in nature.

Because the geological material exhibits uniformity, the expected values of the concentrations,  $E(c_i)$ , and sampling



variances,  $E(s_{c_i}^2)$ , in each of these small samples are the same ( $E(c_i) = E(\bar{c}_{mr})$ ;  $E(s_{c_i}^2) = E(s_{mr}^2)$ ), where the subscript *mr* refers to these small samples. Thus, given that the mean element concentration is:

$$\bar{c}_{mr} = \frac{1}{k} \sum_{i=1}^k c_i, \quad (1)$$

the formula describing the estimation error on this mean element concentration can be derived by propagating the errors in each  $c_i$  through Equation 1 into  $\bar{c}_{mr}$ . This is achieved using a generative formula derived from a Taylor first order expansion about the mean:

$$\sigma_{f(x_i)}^2 \approx \sum_{i=1}^k \left( \frac{\delta f(x_i)}{\delta x_i} \right)^2 \sigma_{x_i}^2, \quad (2)$$

assuming all concentrations are independent (Meyer, 1975; Stanley, 1990). All of the partial derivatives of Equation 1 ( $\delta \bar{c}_{mr} / \delta c_i$ ) required in Equation 2 equal  $1/k$ , and because all  $\sigma_{c_i}^2 = \sigma_{mr}^2$ , the variance on the observed mean becomes:

$$\begin{aligned} \sigma_{\bar{c}_{mr}}^2 &\approx \sum_{i=1}^k \left( \frac{\delta \bar{c}_{mr}}{\delta c_i} \right)^2 \sigma_{c_i}^2 = \sigma_{mr}^2 \sum_{i=1}^k \left( \frac{1}{k} \right)^2 \\ &= \sigma_{mr}^2 \left( \frac{k}{k^2} \right) = \frac{\sigma_{mr}^2}{k}. \end{aligned} \quad (3)$$

Equation 3 is the conventional formula for the square of the standard error on the mean (the standard variance on the mean; Spiegel, 1975). Note that the approximation in Equation 3 becomes an equality when  $\sigma_{mr}^2$  is sufficiently small.

Finally, multiplying both numerator and denominator on the right side of Equation 3 by  $m_r$ , the mass of these small samples, yields:

$$\sigma_{\bar{c}_{mr}}^2 = \frac{\sigma_{mr}^2 m_r}{k m_r} = \frac{\sigma_{mr}^2 m_r}{M_r}, \quad (4)$$

where  $k m_r$  is the combined mass ( $M_r$ ) of the  $k$  small samples taken together.

Now let us consider a single, larger sample with mass  $M$ , as above. The expected value of the concentration of this single large sample [ $E(c_{lv}) = \mu_{lv}$ ] is equal to the expected value of the individual concentrations and mean concentration of the original set of  $k$  small samples [ $E(\bar{c}_{mr}) = E(c_i) = \mu_{mr}$ ]. This is because these two parameters ( $\mu_{Mr}$  and  $\mu_{mr}$ ) are derived from large and small samples of the same homogeneous material. Analogously, the expected value of the sampling variance on the observed element concentration of the large sample concentration [ $E(s_{lv}^2) = \sigma_{lv}^2$ ] will be equal to the expected value of the standard variance on the mean [ $E(s_{mr}^2) = \sigma_{mr}^2$ ] for the original set of  $k$  small samples, again, because these variances are derived from materials that are mechanically equivalent. Given these large and small sample mean and variance equalities, Equation 4 (which describes the col-

lective sample mass-sampling variance relationship for the original  $k$  small samples) can be converted into Equation 5 (which describes the analogous relationship for the large sample):

$$\sigma_{Mr}^2 M_r = \sigma_{mr}^2 m_r \text{ or } \frac{\sigma_{Mr}^2}{m_r} = \frac{\sigma_{mr}^2}{M_r}. \quad (5)$$

Equation 5 indicates theoretically that an inverse relationship exists between sample mass and sampling variance in all homogeneous geological materials. As a result, Equation 5 can be used to determine the sampling variance ( $\sigma_{lv}^2$ ) that will occur in samples of some mass ( $M_r$ ) of a geological material, given the sampling variance ( $\sigma_{mr}^2$ ) observed in samples of a different mass ( $m_r$ ) of that same geological material.

Furthermore, the logarithm of Equation 5 is:

$$\log \sigma_{Mr}^2 = -\log M_r + \log (\sigma_{mr}^2 m_r), \quad (6)$$

Thus, on a logarithmically-scaled diagram of sample mass [ $\log M_r$ ] versus sampling variance [ $\log \sigma_{Mr}^2$ ], Equation 6 describes a straight line with an intercept of  $\log (\sigma_{mr}^2 m_r)$ , and a slope of -1 (Figure 1). The intercept of this line,  $\log (\sigma_{mr}^2 m_r)$ , represents a single fundamental sampling parameter for any geological material under consideration, precisely because it alone establishes the relationship between sample mass and sampling variance for a real geological material. Thus, by plotting the observed sampling variance against the mass of samples used to estimate that variance on such a diagram, one can deduce the value of this fundamental sampling parameter ( $\log (\sigma_{mr}^2 m_r) = \psi_r$ ; which has units of mass; cf. the  $K_s$  of Ingamells and Switzer, 1973) by determining the logarithm of the sampling variance of this line at a unit sample mass (in whatever mass units apply; Figure 1).

Note that the above derivation of the sample mass-sampling variance relationship does not rely on assumptions about the distribution of the sampling errors, nor does it require any simplifying assumptions about the physical characteristics of the sample (e.g., particulate character, size distribution, mineralogy, density, liberation, or shape) other than that it is homogeneous at the scale of the sample size. Furthermore, this derivation is independent of the concentration of the element of interest. Consequently, this inverse relationship between sample mass and sampling variance applies to all sampling errors at all concentrations in all geological materials.

### Equant Grain Model

Historical efforts to estimate sampling error have not generally employed the inverse relationship between sample mass and sampling variance for real geological materials described in Equation 5. Rather, these efforts have typically, and unnecessarily, relied on simplifying assumptions that define an EGM that mimics the sampling char-



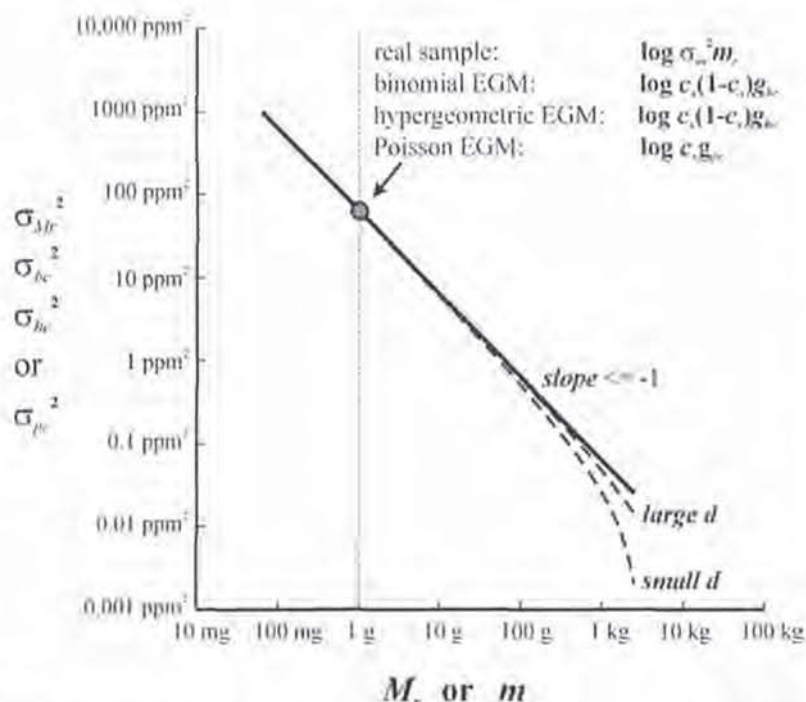


Fig. 1. Logarithmically scaled scatterplot of sample mass ( $m$  or  $M_r$ ) versus sampling variance ( $\sigma^2$ ) for a real sample of any geological material exhibiting any element concentration (real sample mass— $M_r$  vs. real sampling variance— $\sigma^2$ ), and of binomial (sample mass— $m$  vs. binomial sampling variance— $\sigma^2$ ), hypergeometric (sample mass— $m$  vs. hypergeometric sampling variance— $\sigma^2$ ) and Poisson (sample mass— $m$  vs. Poisson sampling variance— $\sigma^2$ ) equant grain models. The solid line describes the limiting sample mass-sampling variance relationship and has a slope of -1. The dashed lines describe the sample mass-sampling variance relationship for hypergeometric EGMs with different  $d$  ( $= N/n$ ; population size/sample size) parameters. The intercept for the real sample line ( $\log \sigma^2 m$ ) is a function of the small real sample mass ( $m$ ) and the small real sampling variance ( $\sigma^2$ ), the product of which is itself a real sampling parameter. The intercepts for the binomial ( $\log c_i(1-c_i)g_{bi}$ ), hypergeometric ( $\log c_i(1-c_i)g_{hi}$ ), and Poisson ( $\log c_i g_{pi}$ ) EGM lines are functions of the concentration of ore grains ( $c_i$ ) and the equant grain masses for the binomial, hypergeometric and Poisson EGMs ( $g_{bi}$ ,  $g_{hi}$ , and  $g_{pi}$ ). These equant grain masses are ideal sampling parameters for each model. All of these sampling parameters (real and ideal) are independent of concentration.

acteristics of the geological material under consideration, to allow application of simple statistical theory to predict sampling error (Herden, 1960; Clifton et al., 1969; Engels and Ingamells, 1970; Ingamells et al., 1972; Ingamells and Switzer, 1973; Ingamells, 1974a,b, 1981, 1986; Ingamells and Engels, 1976; Cheng, 1995; Sketchley, 1997, 1998; Stanley, 1998, 2003). Fortunately, the sample mass-sampling variance relationship in real samples and EGMs based on the binomial distribution are the same (an inverse relationship).

However, the binomial distribution is not the only distribution employed to estimate sampling error. In fact, a number of different approaches involving two other statistical distributions, the Poisson distribution and hypergeometric distribution, have also been used to estimate sampling error (Clifton et al., 1969; Engels and Ingamells, 1970; Ingamells et al., 1972; Ingamells and Engels, 1976; Ingamells, 1981; Stanley, 1998, 2003). These alternative approaches have similarly and necessarily required analogous simplifying

assumptions (i.e., EGMs) so that the properties of these statistical distributions can be employed to estimate sampling error. Conveniently, these alternative approaches employ the same general set of assumptions (EGMs) to allow application of statistical theory to the problem of modeling sampling errors. These assumptions (EGMs) are summarized below.

An EGM is a fictitious, ideal, synthetic geological material meant to exhibit the sampling characteristics of a real geological sample (Herden, 1960; Clifton et al., 1969; Ingamells, 1981; Stanley, 1998), but which has physical characteristics that are consistent with the assumptions required by the statistical distribution used to model the sampling error observed in the real geological material. Consequently, an EGM consists of mineral grains that are all of the same mass and shape. These grains are assumed to be of two, fully liberated, types: ore and gangue. To simplify matters, ore grains are assumed to consist solely of the element or commodity of interest to be analyzed (cf., Ingamells, 1974a,b, 1981, 1986). Thus, although a real Cu ore, say from a sandstone-hosted stratabound Cu deposit, may contain both chalcocite ( $\text{Cu}_2\text{S}$ ; 34.63% Cu) and bornite ( $\text{Cu}_5\text{FeS}_4$ ; 63.31% Cu) ore minerals, as well as a certain small amount of non-recoverable Cu in the gangue minerals, the ore grains in the associated EGM are assumed to consist solely of native copper grains (i.e., 100% Cu). Similarly, gangue grains consist only of the

other elements in the geological material that are not of interest, and are assumed to have a constant composition. Because only one element is considered to be of interest at a time, the composition of the gangue grains is not constrained to be equal to the bulk composition of the other elements in the geological material, and thus can exhibit any composition (i.e., they can be any mineral with any density), provided that they contain none of the element of interest (i.e., 0% Cu). It is commonly convenient to assume that the gangue grains consist of a simple mineral that occurs most abundantly in the geological material under consideration (e.g., quartz in the stratabound Cu deposit example) so that a gangue mineral density can be assigned.

The EGM is constrained to exhibit the same element concentration as the real geological material it is meant to represent. Hence, the concentration (or proportion) of ore grains ( $c$ ) is constrained such that:

$$c = c_o c_o + (1 - c_o) c_g, \quad (7)$$



where  $c$  is the concentration of the element of interest in the sample,  $c_e$  is the concentration of the element of interest in the equant ore grains (100%), and  $c_g$  is the concentration of the element of interest in the equant gangue grains (0%). Under these assumptions, Equation 7 reduces to:

$$v = c_e \quad (8)$$

Obviously, because the equant ore and gangue grains have different compositions, their densities ( $\rho_o$  and  $\rho_g$ , respectively) will differ, so the bulk density of the EGM ( $\rho_e$ ) will be equal to:

$$\rho_e = c_e \rho_o + (1 - c_e) \rho_g \quad (9)$$

Note that because the equant ore and gangue grains consist of ideal minerals that are different from those within the real geological material under consideration, the bulk density of the equant grain model ( $\rho_e$ ) will not necessarily equal the density of the real geological material ( $\rho$ ). Fortunately, this discrepancy has no impact on the ability of the EGM to simulate the sampling characteristics of real geological materials, provided that only the mass of a geological sample, and not its volume, is considered. In addition, if density considerations are important, then the density of the gangue grains can be defined such that the EGM density (at the concentration of interest) equals the density of the real geological material ( $\rho_e$ ) it is meant to represent:

$$\rho_g = \frac{\rho_e - c_e \rho_o}{(1 - c_e)} \quad (10)$$

Furthermore, because the ore and gangue grains have the same mass but have different (assumed) compositions, and thus densities, ore and gangue grains will have different volumes, and thus different dimensions. Assuming these equant grains are spheres, they will have different diameters. However, because mineral densities typically differ by less than an order of magnitude (e.g., in the sandstone-hosted stratabound Cu deposit example, above, the equant native copper ore grains have a density of 8.90 g/cm<sup>3</sup>, whereas the gangue quartz grains have a density of 2.65 g/cm<sup>3</sup>; thus, these densities differ by a factor of 3.68), the volumes of equant ore and gangue grains will also differ by significantly less than an order of magnitude. Furthermore, because the diameter of a spherical equant grain is proportional to the cube root of its volume, the difference between ore and gangue grain diameters will generally be much less than the cube root of 10 (= 2.15; in the example above, the ore and gangue grain diameters will differ by a factor of 1.50). Consequently, although the physical size of equant ore and gangue grains differ in an EGM, the difference is typically small relative to the range of particle sizes observed in real geological materials, which commonly exhibit grain size ranges over several orders of magnitude (e.g., Stanley, 2006a).

Lastly, the size of an EGM sample is typically described by the total number of equant ore and gangue grains in the sample ( $n$ ). Thus, the number of ore ( $x$ ) and gangue ( $n - x$ ) grains equal:

$$x = nc_e \text{ or: } n - x = n(1 - c_e) \quad (11)$$

Because the mass of a sample of an EGM is a function of  $n$  ( $m = ng_e = n v_g \rho_g$ ; where  $g_e$  is the mass of the equant grains and  $v_g$  is the average volume of the equant grains), the mass of an EGM sample can be adjusted to ensure that it exhibits the same sampling characteristics as a real geological material by changing the number of equant grains ( $n$ ) in the EGM sample. In this way, the EGM sample mass can be calibrated such that its sampling error ( $\sigma$ ), or relative sampling error ( $\sigma / \mu = CV$ ), is the same as that observed in a real geological material under consideration (Herden, 1960; Clifton et al., 1969; Stanley, 1998).

Examples of three EGMs associated with different statistical distributions (binomial, hypergeometric, and Poisson) are presented in Figure 2. In each, the large square defines the physical limits of the EGM population, whereas the small square defines the physical limits of an EGM sample. These differ according to the assumptions associated with the statistical distribution employed. Similarly, the proportions (concentrations) of equant ore (gray) and gangue (white) grains vary in these three models, also according to the assumptions associated with these statistical distributions.

### Sampling Error Models

Historical efforts to model sampling error have considered a variety of statistical distributions applied to ideal samples from an EGM (Herden, 1960; Clifton et al., 1969; Ingamells, 1981; Stanley, 1998). In the following discussion, the sample mass-sampling variance relationship is investigated for each of three different statistical models commonly used to describe sampling error in geological materials: the binomial, hypergeometric, and Poisson models. In each case, the relationship between sampling variance and sample mass is derived from the original equations for the variance of binomial, hypergeometric, and Poisson variables. These equations, unlike those for the normal distribution, are functions of either the mean concentration of the element of interest, or the mean number of ore grains in an EGM sample. In all cases, an inverse relationship exists (identical to that illustrated by the above proof for real geological materials, and previously illustrated for a binomial EGM; Wilson, 1964; Kleeman, 1967; Ingamells et al., 1972). This indicates that these theoretical models are consistent with the fundamental relationship between sample mass and sampling variance for real geological materials.

#### Binomial Distribution

Consider a sample of size  $n$  collected from an EGM population of size  $N$ . The expected value of the concentration of the element of interest in this sample  $E(c_e)$  will be  $(= X / N)$ , where  $X$  is the number of ore grains in the population. Because only ore and gangue grains exist and either contain or do not contain the element of interest, the probability of collecting a single ore grain at random from the EGM equals this concentration ( $c_e$ ). However, because a sample from the EGM consists of many grains, some ore and some gangue, many grains must be collected from the



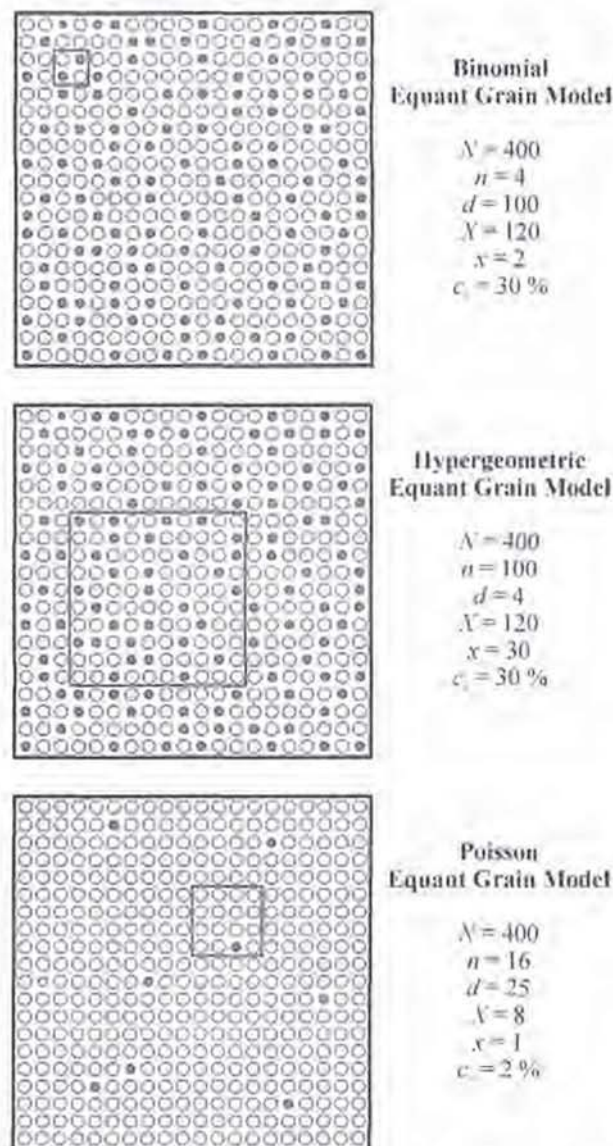


Fig. 2. Schematic examples of various EGMs illustrating the different sampling characteristics of each statistical distribution. Large and small boxes represent the EGM populations and samples, respectively; gray and white circles represent equant ore and gangue grains, respectively.  $N$  = population size;  $n$  = sample size;  $X$  = number of ore grains in the population;  $x$  = number of ore grains in the sample;  $c_s$  = proportion (concentration) of ore grains in the population and the expected number of ore grains in the sample;  $d$  ( $= N/n$ ) can be thought of as the inverse of sampling density.

EGM to obtain a sample. The grains in this sample could be acquired from the EGM by randomly collecting one grain at a time. Although, in reality, geologists would not typically collect the grains in a sample from an EGM in this manner (they would probably scoop out many equant grains at a time from a pile of grains of a particulate material or collect a single or multiple pieces of rock (mineral aggregate) from a non-disaggregated material), statistical considerations make it is worthwhile to consider the sample collected in this way. At each equant grain collection

step, an ore or a gangue grain will be removed from the EGM population and added to the EGM sample, reducing the number of ore or gangue grains remaining in the EGM population by one. As a result, after the first grain is collected (with probability  $c_s$  that it will be an ore grain), the probability of collecting an ore grain will change, at least slightly, to either  $(X-1)/(N-1)$  or  $X/(N-1)$ , depending on whether the initial grain collected was an ore or gangue grain. If the EGM population size ( $N$ ) is large relative to the sample (i.e.,  $N/n \geq 100$ ), initial selection of a gangue grain will not alter this probability much [ $X/(N-1) \approx X/N = c_s$ ]. Furthermore, if the concentration of ore grains is not small (i.e.,  $>0.01\%$ ), initial collection of an ore grain will similarly not change this probability much [ $(X-1)/(N-1) \approx X/N = c_s$ ].

Consequently, with the collection of each grain in the sample, the probability of success (collection of an ore grain) will not change appreciably, provided that  $c_s$  is large and  $N/n$  is large (say,  $>100$ ). Under these provisos, this sampling scenario is equivalent to 'sampling with replacement' (where the probability of success does not change with successive trials, as in flips of a coin). Sampling error associated with this type of sample collection scenario is described by a binomial distribution (Spiegel, 1975).

In geochemical applications, the binomial distribution historically has been used as a sampling model for elements that represent major proportions of a sample (i.e., they typically comprise greater than say 0.01% of the sample mass; Wilson, 1964; Kleeman, 1967; Spiegel, 1975). The sample mass-sampling variance relationship for a binomial variable can be derived using the equation describing the magnitude of sampling variance for a binomial variable:

$$\sigma_b^2 = nc_s(1 - c_s), \quad (12)$$

(Spiegel, 1975) where  $c_s$  is the concentration (proportion) of equant ore grains in the population,  $b$  refers to the binomial distribution, and  $n$  is the number of equant grains contained within a sample. This equation describes the magnitude of sampling variance in terms of the total number of equant grains in a sample and the concentration of the element of interest contained in the ore grains within the population.

Unfortunately, Equation 12 describes the sampling variance of the number of ore grains within a sample, and does not describe the sampling variance of the concentration of ore grains in a sample. To obtain an equation that describes the sampling variance for the concentration of ore grains, we must rely on the fact that the concentration (or proportion) of ore grains in the sample approximates the concentration of ore grains in the population ( $c_s$ ), and equals the number of ore grains in a sample ( $x$ ) divided by the total number of grains in the sample ( $n$ ):

$$c_s = \frac{x}{n}. \quad (13)$$

Analogous relationships exist between both the mean and standard deviation of the number of ore grains, and the mean and standard deviation of the concentration of ore



grains in a sample:

$$\mu_c = \frac{\mu_s}{n}, \quad (14)$$

and:

$$\sigma_c = \frac{\sigma_s}{n}. \quad (15)$$

Squaring Equation 15 yields:

$$\sigma_{bc}^2 = \frac{\sigma_{bs}^2}{n^2}. \quad (16)$$

Thus, by dividing both sides of Equation 12 by  $n^2$ , we obtain a binomial sampling variance equation in terms of the concentration of ore grains ( $\sigma_{bc}^2$ ), instead of in terms of the number of ore grains ( $\sigma_{bs}^2$ ):

$$\sigma_{bc}^2 = \frac{c_s(1-c_s)}{n}. \quad (17)$$

Taking the square root of Equation 17 yields the form of the sampling standard deviation as a function of the mean concentration:

$$\sigma_{bc} = \sqrt{\frac{c_s(1-c_s)}{n}}, \quad (18)$$

and this relationship is illustrated in Figure 3. Note that binomial sampling error reaches a maximum at concentrations of 50%, and equals zero at concentrations of 0% and 100%. Furthermore, the larger the sample ( $n$ ), the smaller the binomial sampling error ( $\sigma_{bc}$ ).

Finally, multiplying the numerator and denominator of the right side of Equation 17 by  $g_{bc}$ , the mass of the equant grains in a sample from a binomial EGM, we obtain:

$$\sigma_{bc}^2 = \frac{c_s(1-c_s)g_{bc}}{ng_{bc}} = \frac{c_s(1-c_s)g_{bc}}{m}, \quad (19)$$

where  $n \times g_{bc} = m$ . Thus, an inverse relationship exists between the sample mass ( $m$ ) and the binomial sampling variance for that sample ( $\sigma_{bc}^2$ ).

Taking the logarithm of Equation 19 yields:

$$\log \sigma_{bc}^2 = -\log m + \log(c_s(1-c_s)g_{bc}). \quad (20)$$

Because the logarithm of  $m$  is the abscissa and the logarithm of  $\sigma_{bc}^2$  is the ordinate on a logarithmically scaled scatterplot of sample mass versus sampling variance, Equation 20 is the formula for a straight line with a slope of -1 and intercept of  $\log(c_s(1-c_s)g_{bc})$  on Figure 1. As above, the intercept of this line with the vertical axis where  $m = 1$  [ $\log(c_s(1-c_s)g_{bc}) = \psi$ ] is a sampling parameter for the binomial EGM under consideration, precisely because it establishes the relationship between sample mass and binomial sampling variance (as in Equation 19). However, unlike that demonstrated above for real geological materials, this sampling parameter is a function of concentration. Thus, a more fundamental sampling parameter is  $g_{bc}$ , the mass of each equant grain in the binomial EGM model. By means analogous to that for real geological materials,

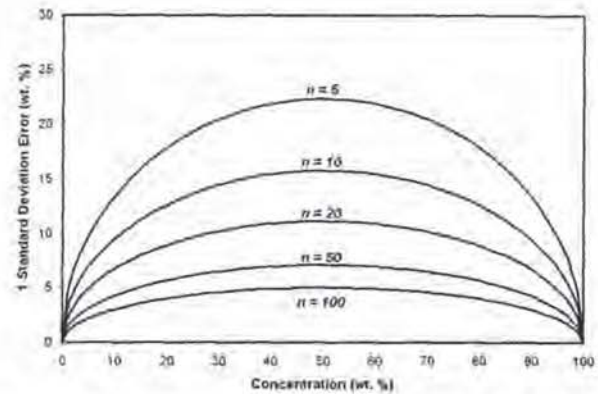


Fig. 3. Examples of binomial sampling error models with different total numbers of equant grains ( $n$ ). Maximum sampling error occurs at a concentration of 50%, and is symmetric across the interval 0%  $\leftrightarrow$  100%. Larger  $n$  define binomial sampling error models with smaller errors, as indicated.

above, one can deduce the magnitude of this fundamental binomial sampling parameter ( $g_{bc}$ ) by determining the logarithm of the sampling variance of this binomial sampling line at a unit sample mass for a given concentration, or:

$$g_{bc} = \frac{\sigma_{bc}^2 m}{c_s(1-c_s)}. \quad (21)$$

#### Hypergeometric Distribution

Like the binomial distribution, the hypergeometric distribution is also used as a sampling model when the element of interest is a major component of a sample (i.e., it typically comprises greater than 0.01% of the mass; Stanley, 2003). However, unlike the binomial distribution, this distribution is used to model sampling error when the size of the sample ( $n$ ) is large relative to the size of the population ( $N$ ) from which it is drawn (typically, when  $N/n < 100$ ). Scenarios where this distribution is likely to be important in geological applications include when drill core or drill cutting samples are collected, where the mass of the sample (half core or a split portion of the cuttings) is generally 1/2 to 1/10<sup>th</sup> the mass of the population from which it is derived (the whole core or all of the cuttings).

The hypergeometric distribution bears a strong similarity to the binomial distribution (Spiegel, 1975). However, an important difference is that whereas the probability of success in successive trials under a binomial sampling model (collection of an ore grain) does not change, under a hypergeometric sampling model it does. For example, suppose that the total number of equant grains ( $N$ ) in an EGM population is 100 000 and that the concentration of the element of interest ( $c_s$ ) is 5%: the number of equant ore grains ( $X$ ) in the population is then 5000. If a sample of the EGM has 20 grains (thus,  $N/n = 5000$ ), then the probability of collecting the 20<sup>th</sup> (last) equant grain for the sample will range from  $(X-19)/(N-19) [= 4981/99\,981 = 4.98195\%]$  to  $X/(N-19) [= 5000/99\,981 = 5.00095\%]$  depending on how many of the first 19 randomly collected equant grains



are ore grains. Both of these end-member probabilities are similar to the original probability of 5% (differing by less than 1%, relative, from the initial probability). Thus, the probability of collection of an ore grain for each of the 20 trials will be essentially the same (i.e., this scenario is virtually equivalent to 'sampling with replacement', a binomial process).

However, if the EGM has the same element concentration, but is smaller, with say  $N = 1000$  equant grains, then  $X = 50$ . If we collect a sample of identical size from this smaller EGM,  $n$  still equals 20, and thus  $N/n = 50$ . Consequently, the probability of collecting an ore grain as the last equant grain will again range from  $(X - 19) / (N - 19) [= 31/981 = 3.16004\%]$  to  $X / (N - 19) [= 50/981 = 5.09684\%]$ , again depending on how many of the first 19 randomly collected grains are ore grains. However, in contrast to the first scenario where  $N/n = 5000$ , these probabilities are very different from the original probability of 5% (differing by almost 40%, relative, from the initial probability). Because the probability of success for successive trials gradually changes in this scenario, these sampling errors are described using a hypergeometric distribution.

Thus, for samples where the element concentration is significant (say,  $>0.01$  wt.%) and the sample mass is a large proportion of the population from which it is collected (i.e., drill core and cuttings), the hypergeometric distribution should be used as a model to describe sampling error.

The sample mass–sampling variance relationship for a hypergeometric variable can be derived using the equation describing the magnitude of sampling variance for a hypergeometric variable:

$$\sigma_m^2 = nc_s(1-c_s) \left( \frac{N-n}{N-1} \right), \quad (22)$$

(Spiegel, 1975), where  $n$ ,  $N$ ,  $c_s$ , and  $x$  are as above, and  $h$  refers to the hypergeometric distribution. Note that Equation 22 is very similar to the sampling variance equation for a binomial variable (Equation 12), differing only by a trailing factor that is a function of the size of the sample ( $n$ ) and the size of the population ( $N$ ) from which it is drawn. Equation 22 describes the magnitude of sampling variance in terms of the total number of equant grains in a sample ( $n$ ), the total number of equant grains in a population ( $N$ ), and the concentration (proportion;  $c_s$ ) of ore grains in the sample.

As with the binomial model, and for the same reasons above, dividing both sides of Equation 22 by  $n^2$  yields:

$$\sigma_{hc}^2 = \frac{c_s(1-c_s)}{n} \left( \frac{N-n}{N-1} \right), \quad (23)$$

Furthermore, dividing the numerator and denominator of the trailing factor in Equation 23 by  $n$  and defining  $N/n$  as the population size/sample size ratio ( $d$ ), a known or easily approximated quantity in most sampling scenar-

ios, we obtain:

$$\begin{aligned} \sigma_{hc}^2 &= \frac{c_s(1-c_s)}{n} \left( \frac{\frac{N}{n} - \frac{n}{n}}{\frac{N}{n} - \frac{1}{n}} \right) \\ &= \frac{c_s(1-c_s)}{n} \left( \frac{d-1}{d-1/n} \right), \end{aligned} \quad (24)$$

Then, multiplying the numerator and denominator of the right side of Equation 24 by  $g_{hc}$ , the mass of each equant grain in the hypergeometric EGM, yields:

$$\begin{aligned} \sigma_{hc}^2 &= \frac{c_s(1-c_s)g_{hc}}{ng_{hc}} \left( \frac{d-1}{d-1/n} \right) \\ &= \frac{c_s(1-c_s)g_{hc}}{m} \left( \frac{d-1}{d-g_{hc}/m} \right), \end{aligned} \quad (25)$$

because  $n = m / g_{hc}$ .

Taking the square root of Equation 25 yields the functional form of sampling standard deviation as a function of the mean:

$$\sigma_{hc} = \sqrt{\frac{c_s(1-c_s)g_{hc}(d-1)}{m(d-g_{hc}/m)}}, \quad (26)$$

and this relationship is illustrated in Figure 4. As with binomial sampling error, hypergeometric sampling error reaches a maximum at a concentration of 50%, equals zero at 0% and 100% concentrations, and is smaller with larger samples (large  $n$ ). Furthermore, as  $N/n$  decreases (the sample becomes large relative to the population), the hypergeometric sampling error decreases, precisely because the sample becomes more representative of the population.

Equation 25 now describes the magnitude of hypergeometric sampling error associated with the concentration of the element of interest in the sample in terms of sample mass ( $m$ ). However, in contrast to the binomial model, above, an exact inverse relationship does not exist between sample mass ( $m$ ) and the variance of sampling error in that sample ( $\sigma^2$ ).

Nevertheless, if we take the logarithm of Equation 25, we obtain:

$$\begin{aligned} \log \sigma_{hc}^2 &= \left[ -\log m + \log \left( \frac{d-1}{d-g_{hc}/m} \right) \right] \\ &\quad + \log (c_s(1-c_s)g_{hc}). \end{aligned} \quad (27)$$

In Equation 27, both terms in the square brackets are a functions of  $m$ , and thus define the slope of a (curved) line on a logarithmic diagram plotting sample mass  $[\log m]$  versus sampling variance  $[\log \sigma_{hc}^2]$ . The last term in Equation 27 defines the intercept  $[\log (c_s(1-c_s)g_{hc})]$  of this line (a grade dependent sampling parameter  $\psi_s$ ).

Furthermore, because the average mass of the equant grain ( $g_{hc}$ ) is constant, as the sample mass ( $m$ ) decreases



to  $g_{he}$  (its minimum limit, corresponding to a sample composed of one equant grain), the ratio in the second term in the square brackets of Equation 27 progressively converges on unity. As a result, the logarithm of this ratio converges on zero, and the slope of the line at small sample masses converges on  $-1$ , regardless of the magnitude of  $d$ . In contrast, with large sample masses,  $g_{he}/m$  converges on zero, and the ratio in the second term in the square brackets of Equation 27 progressively converges on  $(d-1)/d$ . When  $d$  is large, this ratio is nearly 1, so its logarithm is nearly zero, and the slope of the line remains equal to  $-1$ , or nearly so, for all sample masses. Consequently, when  $d$  is large, Equation 27 defines a line for the hypergeometric EGM that is essentially identical to the line defined by Equation 20 for the binomial EGM (Figure 1). However, when  $d$  is small (e.g., equal to two, which corresponds to the scenario when samples of half drill core are collected from full drill core), this ratio becomes substantially less than 1, and so the logarithm of the ratio containing  $d$  in Equation 27 becomes substantially less than 0 (e.g., for  $d=2$ ,  $\log(d-1)/d = -0.30103$ ). Thus, for small  $d$ , Equation 27 defines a curve that becomes asymptotic to a line with slope of  $-1$  for small sample masses, but which gradually steepens at large sample masses.

The net effect of the second term in the square brackets of Equation 27 is thus to bend the hypergeometric sampling line downward as sample mass increases, away from the straight binomial sampling line, so that a lower sampling variance will be predicted for a given large sample mass (Figure 1). This smaller sampling variance is an intuitive consequence of collecting a sample which is large relative to the population from which it is derived (i.e., because a larger sample is more representative of its population, it should exhibit a smaller sampling error).

Lastly, as with the binomial EGM,  $c_x(1-c_x)g_{he} = \psi_h$  is a grade-dependent sampling parameter, and  $g_{he}$  is a more fundamental sampling parameter for the hypergeometric EGM. This more fundamental sampling parameter can be determined by:

$$g_{he} = \frac{\sigma_{ps}^2 md}{c_x(1-c_x)(d-1) + \sigma_{ps}^2}. \quad (28)$$

#### Poisson Distribution

Like the binomial and hypergeometric distributions, the Poisson distribution is also used as a sampling model (Wickman, 1962; Clifton et al., 1969; Ingamells, 1974a,b, 1981; Stanley, 1998, 2003, 2006a). However, this distribution applies to scenarios where the element of interest is only a minor component of a sample (i.e., it typically comprises less than, say, 0.01% of the sample mass). Furthermore, this distribution applies when the element of interest occurs only in a small number of ore grains (generally  $<50$ ) within the sample (referred to as nuggets). Consequently, the Poisson distribution historically has been used in the estimation of sampling errors for Au, PGE, diamonds, etc. (Clifton et al., 1969; Stanley, 1998, 2003, 2006a,b).

Like the binomial and hypergeometric sampling scenar-

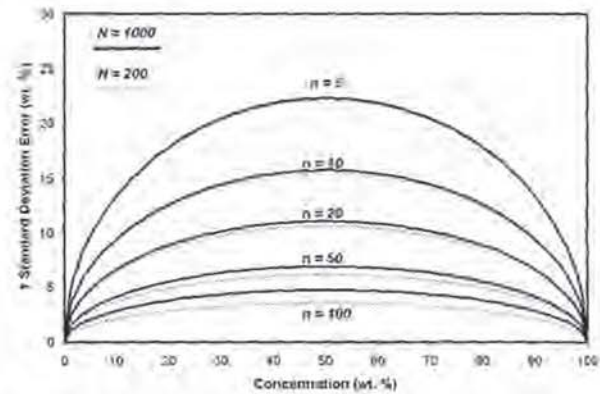


Fig. 4. Examples of hypergeometric sampling error models with different sampling parameters ( $n$ ) for two different population sizes ( $N=200$  and  $N=1000$ ). In all cases, the maximum sampling error occurs at a concentration of 50%, and is also symmetric across the interval 0%–100%. Larger  $n$  define hypergeometric sampling error models with smaller errors; however, as  $n$  (the size of the sample) approaches  $N$  (the size of the population from which the sample is drawn), the magnitude of sampling error decreases, because the sample becomes more representative of (larger relative to) the population.

ios above, the sample mass–sampling variance relationship for Poisson variables can be derived using the equation describing the magnitude of sampling variance for a Poisson variable:

$$\sigma_{ps}^2 = x, \quad (29)$$

(Spiegel, 1975) where  $x$  is the number of rare ore grains (nuggets) collected in a sample, and  $\sigma_{ps}^2$  is the Poisson sampling variance for this small number of nuggets.

Unfortunately, geochemical analyses do not typically report the number of nuggets in a sample, but rather report the concentration of the element contained in nuggets within a sample (e.g., the gold concentration). Thus, Equation 29 cannot be used to describe the sampling error of element concentrations in nugget-bearing samples because it does not relate sampling error to concentration. However, if one manipulates Equation 29 by inverting it and then multiplying both sides by the standard deviation ( $\sigma_{ps}$ ), we obtain:

$$\frac{\sigma_{ps}}{x} = \frac{1}{\sigma_{ps}} = \frac{1}{\sqrt{x}}. \quad (30)$$

The characteristics of the EGM require that the relative sampling error on the number of nuggets in a sample ( $\sigma_{ps}/x$ ) from an EGM must equal the relative sampling error on the concentration of the element of interest contained in those nuggets ( $\sigma_{ps}/c$ ; note that both of these parameters are unitless; Clifton et al., 1969; Stanley, 1998, 2003; Stanley and Lawie, in press):

$$\frac{\sigma_{ps}}{x} = \frac{\sigma_{ps}}{c_x}. \quad (31)$$

Hence, the relative error of replicate concentrations of the nugget-borne element can be used to determine the effective number of nuggets ( $x$ ) in an EGM sample (Clifton



et al., 1969; Stanley, 1998, 2003):

$$\frac{\sigma_{pc}}{c_s} = \frac{1}{\sqrt{x}} \text{ or } \frac{\sigma_{pc}^2}{c_s^2} = \frac{1}{x}. \quad (32)$$

After some algebraic manipulation, and substituting  $n \times c_s$  for  $x$  in Equation 32, we obtain:

$$\sigma_{pc}^2 = \frac{c_s^2}{x} = \left( \frac{c_s}{n} \right) c_s = k_s c_s = \frac{c_s^2}{nc_s} = \frac{c_s}{n}, \quad (33)$$

where  $n$  is the total number of equant grains in a sample,  $k_s$  is the concentration contribution per ore grain ( $= c_s / x$ ), and  $c_s$  is the proportion of equant ore grains in the sample. Multiplying the numerator and denominator of the right side of Equation 33 by  $g_{pc}$ , the mass of each equant grain in a sample from the Poisson EGM, yields:

$$\sigma_{pc}^2 = \frac{g_{pc} c_s}{ng_{pc}} = \frac{g_{pc} c_s}{m}, \quad (34)$$

which expresses the sampling variance as an inverse function of sample mass ( $m$ ). The square root of Equation 34 describes the Poisson sampling variance as a function of concentration on a mean versus standard deviation scatter-plot:

$$\sigma_{pc} = \sqrt{\frac{g_{pc} c_s}{m}} \quad (35)$$

(Figure 5).

Taking the logarithm of Equation 34 yields:

$$\log \sigma_{pc}^2 = -\log m + \log(g_{pc} c_s), \quad (36)$$

which is an equation for a straight line with slope of  $-1$  and intercept of  $\log(g_{pc} c_s)$  on a logarithmic diagram of sample mass ( $m$ ) versus sampling variance ( $\sigma_{pc}^2$ ; Figure 1).

Thus, as  $c_s$  approaches zero, the  $(1 - c_s)$  factors in the intercept terms of Equations 20 and 27 converge on unity.

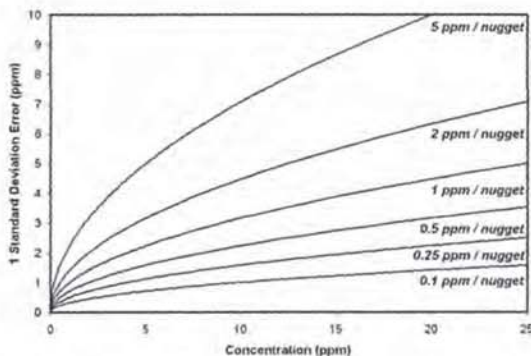


Fig. 5. Examples of Poisson sampling measurement-error models with different size nuggets that contribute different concentrations of gold per grain ( $k_s$ ) to each sample. The coarser the gold, the larger the nugget effect and the larger the sampling error at a given concentration.

Therefore, at small concentrations, the intercepts for the lines describing the relationship between sample mass and sampling variance in binomial, hypergeometric, and Poisson EGMs converge (on the intercept term of Equation 36). Because the slopes and intercepts of these lines are equal at the limits, all of these lines on logarithmic diagrams of sample size versus sampling variance (for binomial, hypergeometric, and Poisson sampling variances; Figure 1) converge to the same line for small samples and low concentrations. This illustrates that these three distributions complement each other in different sampling scenarios. Furthermore, these distributions provide equivalent estimates of sampling variance at the limits of their application.

Finally, as above,  $g_{pc} c_s = \psi_p$  is a grade-dependent sampling parameter for the Poisson EGM, but  $g_{pc}$  is a more fundamental sampling parameter that can be determined by:

$$g_{pc} = \frac{\sigma_{pc}^2 m}{c_s}. \quad (37)$$

## Discussion

The above theoretical derivations provide an important

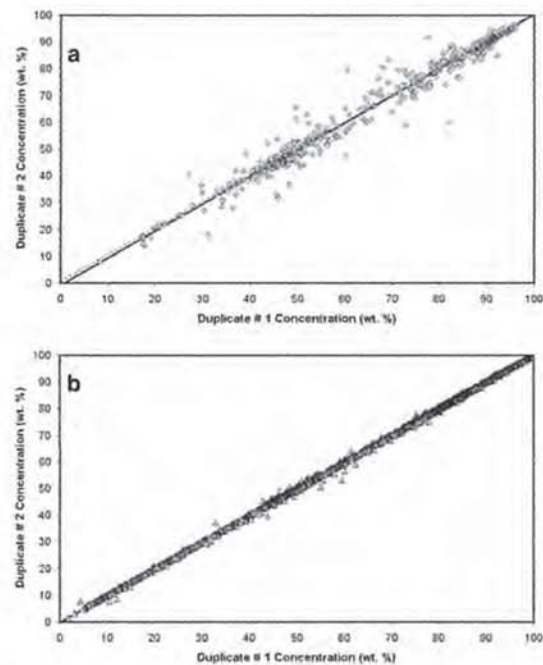


Fig. 6. Scatterplots of *a.* 522 rig duplicate and *b.* 2112 post-crushing duplicate  $\text{Fe}_2\text{O}_3$  assays from a BIF Fe deposit. Dashed lines are  $X = Y$ ; regression lines (solid) that assume error in both variables proportional to the standard deviations (standard deviation lines) have slopes and intercepts of 1.0119 and  $-0.8425$  for *a.*, and 1.0014 and  $-0.1088$  for *b.*, respectively. Correlation coefficients are 0.9865 and 0.9988, respectively.



foundation for the examination and quantification of sampling variance in ores. However, more importantly, these results provide a substantial and powerful mass-standardization tool with which to analyze sampling variances in a variety of mineral deposits and mineral deposit types using the three statistical distributions considered above. Below, two example duplicate data sets from an anonymous banded iron formation (BIF) Fe deposit are examined and evaluated to illustrate how mass-standardization of sampling errors can be used to obtain comparable results.

The example data sets under consideration consist of  $\text{Fe}_2\text{O}_3$  concentrations determined from relatively dry duplicate drill chip samples collected using a single reverse circulation drill rig (Figure 6). The rocks sampled consist of different proportions of BIF (ideally, a sub-equal mixture of quartz and hematite;  $\rho \approx 3.96$ ), iron ore (ideally, massive hematite;  $\rho \approx 5.27$ ), and interbedded shale ( $\rho \approx 2.65$ ). Drill chip samples were derived from 3 m long sampling intervals of 5.5" diameter holes; the drill chip fragments were all less than 3 cm in diameter, and approximately 95% of these passed through a 1.2 cm sieve. The total masses of the drill chips available for sampling ranged from approximately 120 kg (for shale) to 180 kg (for BIF) to 240 kg (for iron ore), depending on the composition and density of the rocks.

This material was split at the rig drill using a rotary cone splitter. Rig samples collected on site averaged approximately 6 kg. Sample preparation involved crushing the entire sample to  $-2.8$  mm, and then sub-sampling one quarter of this material (approximately 1.5 kg) for pulverization to  $-105$   $\mu\text{m}$ . After pulverization, a 5 g subsample analyte was fused and determined for  $\text{Fe}_2\text{O}_3$  by conventional XRF analysis.

During this sampling, preparation, and analysis protocol, 522 pairs of rig duplicates, and 2112 pairs of post-crushing duplicates were collected. The rig duplicates exhibit an average mass of 6.147 kg (ranging from 0.338 to 18.100 kg), whereas the post-crushing duplicates exhibit an average mass of 1.314 kg (ranging from 0.106 to 4.350 kg).

Because each of these rig and post-crushing duplicate samples exhibited different masses (Figure 7), calculation

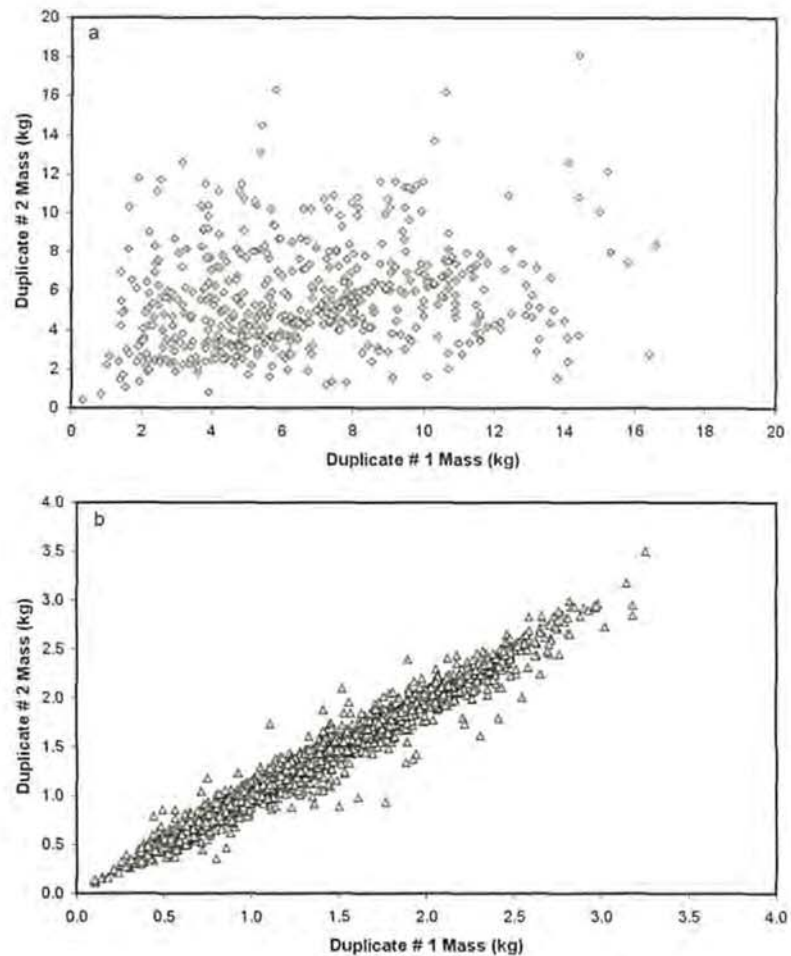


Fig. 7. Scatterplot of the masses of *a*. 522 rig duplicate samples and *b*. 2112 post-crushing duplicate samples from a BIF deposit.

of the duplicate means and standard deviations could not be undertaken by conventional means. Instead, weighted duplicate means ( $\bar{c}_w$ ) and standard deviations ( $s_w$ ) were calculated for each type of duplicate pair using the following formulae:

$$\bar{c}_w = \frac{\sum_{i=1}^2 m_i c_i}{\sum_{i=1}^2 m_i} \quad (38)$$

and:

$$s_w = \sqrt{\frac{\sum_{i=1}^2 m_i (c_i - \bar{c}_w)^2}{\frac{1}{2} \sum_{i=1}^2 m_i}} \quad (39)$$



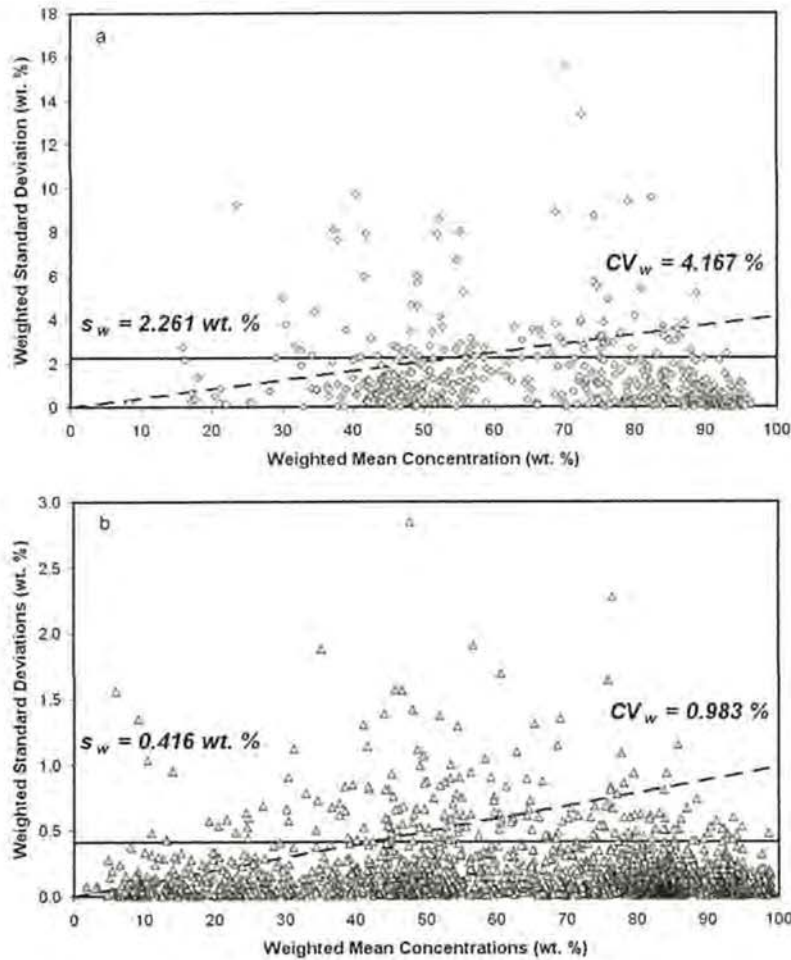


Fig. 8. Thompson-Howarth scatterplots of weighted mean concentrations versus weighted standard deviations for *a*, 522 rig duplicate samples and *b*, 2112 post-crushing duplicate samples from a BIF deposit. The solid horizontal lines represent the average (RMS) weighted standard deviation of the duplicate pairs, and the dashed sloping lines through the origin represent the average (RMS) relative error ( $\mu/\sigma$ ) of the duplicate pairs.

where  $m_i$  are the duplicate sample masses, and  $c_i$  are the duplicate concentrations. These statistics for both duplicate data sets are plotted in Figure 8. Furthermore, because the average mass of each duplicate pair also differed, corrections to the observed standard deviations ( $s_w$ ) were made using Equation 5 and a sample mass datum equal to the mean sample mass observed in each set of duplicate pairs (6.147 kg and 1.314 kg, respectively). This provided estimates of the mass-standardized standard deviations ( $s_{wm}$ ). Finally, average standard deviations and relative errors ( $s/\bar{x}$ ) for each set of duplicates were calculated. Again, this can not be undertaken by conventional means because variances, not standard deviations, are additive, and use of the conventional mean formula (cf. Equation 1) does not produce an un-biased estimate of the mean standard deviation or relative error. The unbiased root mean square

(RMS) standard deviations ( $s_{wm,RMS}$ ) of 2.261 wt.% and 0.416 wt.%, and un-biased RMS relative errors ( $s_{wm}/\bar{c}_{wm,RMS}$ ) of 4.167% and 0.983%, were calculated from these mass-standardized standard deviations for each set of duplicates using the following formulae:

$$s_{wm,RMS} = \sqrt{\frac{\sum_{i=1}^k \bar{g}_i s_{wm,i}^2}{\sum_{i=1}^k \bar{g}_i}} \quad (40)$$

and:

$$\frac{s_{wm}}{\bar{c}_{wm,RMS}} = \sqrt{\frac{\sum_{i=1}^k \bar{g}_i \left( \frac{s_{wm,i}}{\bar{c}_{wm,i}} \right)^2}{\sum_{i=1}^k \bar{g}_i}} \quad (41)$$

where  $s_{wm,i}$  is the weighted, mass-standardized duplicate standard deviation, and  $s_{wm}/\bar{c}_{wm}$  is the weighted, mass-standardized duplicate relative error.

The cumulative frequency distributions of these relative errors are presented in Figure 9 and compared with the frequency distributions of relative errors at different levels that would be expected if the errors were normally distributed. Clearly, the mass-standardized relative errors are distinctly non-normal in both data sets, because the distributions of relative errors do not match the expected distributions for normally distributed relative errors at any relative error level. For both

duplicate data sets, the observed relative error distributions start below what would be the normally distributed error line for the average error observed (4.167%, ~5%, and 0.983%, ~1%) but then climb above these lines at high percentiles. This indicates that the distributions of relative errors in these data sets are more positively skewed than what would be expected from normally distributed relative errors. Thus, the underlying distributions of error are also skewed (and thus not normally distributed). Because of their non-zero skewness, the distributions of sampling error in these data sets bear at least some resemblance to binomial, hypergeometric, and Poisson sampling error, all of which exhibit sampling error frequency distributions that are positively skewed at these concentrations.

These duplicate samples allow determination of both the total and post-crushing measurement errors. Subtraction of



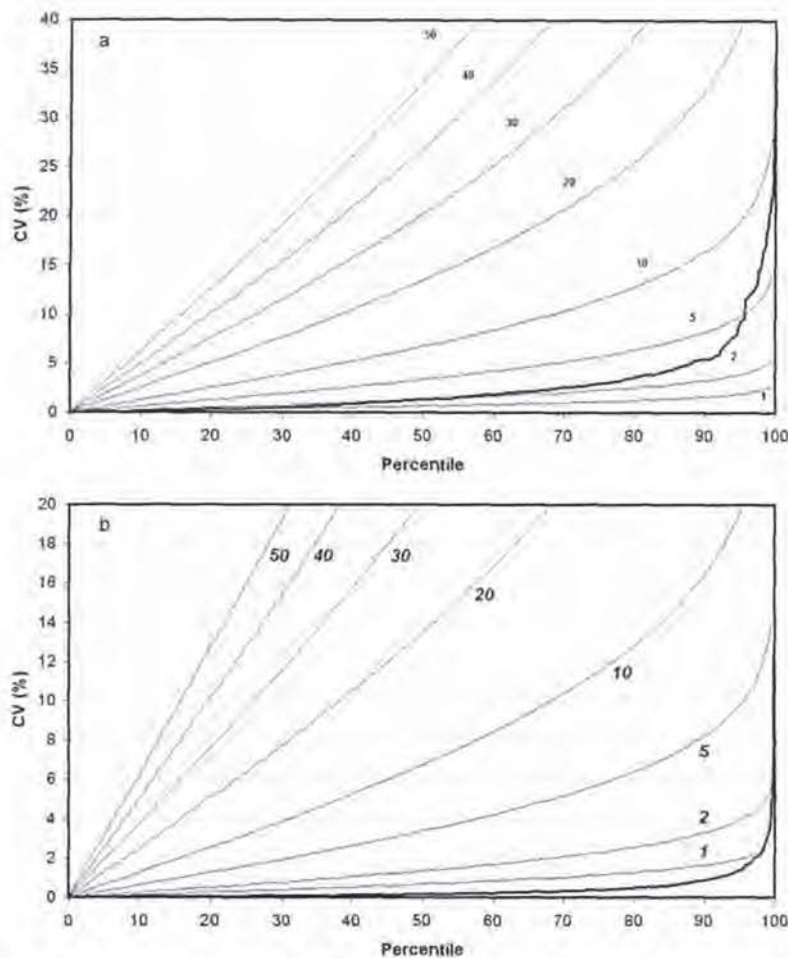


Fig. 9. Percentile plots of the weighted RMS relative errors ( $\mu/\sigma$ ). Thin lines represent expected curves for normally distributed relative errors at various relative error levels (as indicated). Thick lines represent the distribution of relative errors from *a.* 522 rig duplicate samples and *b.* 2112 post-crushing duplicate samples from a BIF deposit. These lines have trends different from the normal curves, indicating that the sampling errors in this deposit are distinctly non-normal.

the post-crushing measurement error variance ( $\sigma_2^2$ ) from the total measurement variance ( $\sigma_1^2$ ) allows determination of the initial sampling variance ( $\sigma_s^2$ ) by the equation:

$$\sigma_s^2 = \sigma_1^2 - \sigma_2^2. \quad (42)$$

Using the observed one standard deviation measurement errors calculated from the rig and post-crushing duplicate data sets, the sampling error ( $\sigma_s$ ) from this BIF deposit is calculated to be 2.222 wt.%. The relative sampling error ( $\gamma_s$ ) can be determined using an analogous equation (Stanley and Smee, in press; = 4.049%). These absolute and relative initial sampling error magnitudes are similar to the total measurement absolute and relative measurement error magnitudes observed in the rig duplicate data (2.261 wt.% and 4.167%, respectively). Thus, sampling error represents approximately 97% and 94% of the absolute and relative variance in the data, indicating that the majority

of measurement error introduced to these samples occurred during sampling with the rotary splitter at the drill rig. Fortunately, the magnitude of this sampling error is relatively low (~2 wt.% absolute and ~4% relative), and would generally not be considered a significant impediment in exploration or resource definition of a BIF deposit, at least provided that this error is not caused by inappropriate sampling and analysis procedures.

The weighted average sample masses and weighted average, mass-standardized, duplicate variances for each data set have been plotted on Figure 10, a logarithmic diagram of sample mass versus sampling variance. This diagram is analogous to that in Figure 1. The weighted average, mass-standardized sampling variance of these duplicates (4.937 wt.%<sup>2</sup>, equal to the square of 2.222 wt.%) has been plotted at the average sample mass to define a line that can be used to estimate the sampling variance that would occur in samples of any sample mass from this BIF deposit. For example, samples of 1 kg mass can be expected to exhibit average sampling variances of 30.351 wt.%<sup>2</sup> ( $\sigma = 5.509$  wt.%, as derived using Equation 5; Figure 10, circle). Similarly, analysis of the entire mass of drill cuttings of different compositions ( $m = 120$  kg for shale,  $m = 180$  kg for BIF, and  $m = 240$  kg for iron ore) can be expected to produce average sampling errors of 0.253 wt.%<sup>2</sup>, 0.169 wt.%<sup>2</sup>, and 0.126 wt.%<sup>2</sup> ( $\sigma = 0.503$  wt.%, 0.411 wt.%, and 0.356 wt.%, Figure 10, triangles), respectively.

Thus, we can predict the expected sampling variance that would occur for all sample masses, from samples as small as (or smaller than) 1 kg to samples consisting of the entire mass of drill cuttings from shale, BIF, or iron ore (Figure 10).

Note that Figure 10b presents sampling lines describing the initial and post-crushing measurement errors for drill chip material from this BIF deposit. These lines are parallel, but have different intercepts, even though the same geological material is contained in these samples. The difference in these intercepts exists because the materials represented by these different lines have different grain sizes. The post-crushing material exhibits lower variance even though smaller samples were collected (contrary to the above derived sample mass-sampling variance relationship). This is because the reduction in grain size caused by crushing resulted in the creation of many more particles in the post-crushing duplicate samples, even though these



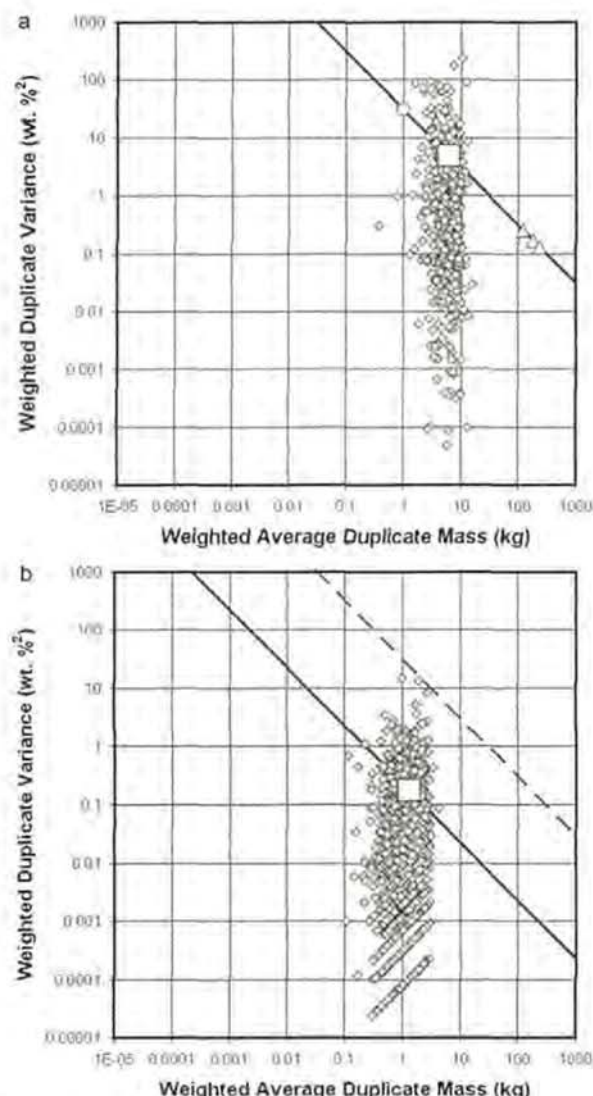


Fig. 10. Logarithmic sample mass versus sampling variance scatterplots for *a*. 522 rig duplicate samples and *b*. 2112 post-crushing duplicate samples from a BIF deposit. The solid lines have slopes of  $-1$  and pass through the large white squares, defining the weighted average duplicate variances at the weighted average sample masses. In *a*, the white circle and white triangles identify the expected measurement error variances for sample masses of 1 kg, 120 kg, 180 kg, and 240 kg, respectively. In *b*, the dashed line is identical to the solid line in *a*.

samples are of smaller size. Consequently, in order to be truly valid, any comparison of the sampling variance of two geological materials must be undertaken using both mass-standardization and a common geological material particle size distribution.

Knowledge of what sampling variance can be expected if larger or smaller samples of a real geological material are collected and analyzed allows cost and effort optimization of sampling procedures (cf., Stanley and Smee, in press). This can result in important cost savings or resource definition improvements during mineral exploration and mining.

## Conclusions

Historical applications of sampling theory have relied on the assumption that sampling variance is inversely proportional to sample mass. This assumption was based on the sampling characteristics of an ideal, physical model (the equant grain model) coupled with a statistical distribution (the binomial distribution). Unfortunately, to date, it was not proven that this ideal model accurately describes the sampling characteristics of real geological materials that exhibit extreme complexity in terms of grain size, shape, composition, and degree of liberation variation.

Proof that sampling variance is inversely proportional to sample mass in real geological materials demonstrates that the foregoing assumption is correct. Furthermore, this same inverse relationship is demonstrated for two other ideal, coupled physical and statistical models (hypergeometric and Poisson equant grain models). Thus, geoscientists can now confidently perform mass-standardization corrections to sampling variance data to allow calculation of average sampling variance from samples with different masses. Furthermore, geoscientists can now confidently compare mass-standardized sampling variances acquired using different sampling methods to assess whether real differences exist in the resulting sampling variances.

Finally, fundamental sampling parameters can be determined for any geological material using binomial, hypergeometric, or Poisson sampling theory assumptions. These parameters are independent of concentration, and summarize the sampling characteristics of different geological materials. Hence, these parameters provide a basis for comparison of the sampling characteristics of different geological materials.

## Acknowledgments

The author would like to acknowledge BHP Billiton, specifically Drs. G. Booth and C. Salt, for providing the BIF deposit duplicate data sets, and for permission to publish this analysis of their quality assurance/quality control data. This paper was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant, and a financial stipend and logistical support from CRC-LEME (Perth, Western Australia) to the author.

## References

- Bettenay, L., and Stanley, C.R., 2001, Geochemical data quality: The "fit-for-purpose" approach: *Explore, Newsletter of the Association of Exploration Geochemists*, v. 11, p. 12, 21–22.
- Cheng, X., 1995, Mass changes during hydrothermal alteration, Silver Queen epithermal deposit, Owen Lake, central British Columbia: Ph.D. thesis, Vancouver, Canada, University of British Columbia, 195 p.
- Clifton, H.E., Hunter, R.E., Swanson, F.J., and Phillips, R.L., 1969, Sample size and meaningful gold analysis: U.S.G.S. Professional Paper 625-C, p. C1–C17.



- Duncan, A.J., 1971, Comments on "A general theory of sampling": *Materials Research and Standards*, v. 11, p. 25.
- Engels, J.C., and Ingamells, C.O., 1970, Effect of sample inhomogeneity in K-Ar dating: *Geochimica et Cosmochimica Acta*, v. 34, p. 1007-1017.
- Gy, P.M., 1982, *Sampling of particulate materials, theory and practice*: New York, Elsevier, 429 p.
- Herden, G., 1960, *Small particle statistics*, 2nd edition: London, Butterworths Scientific Publications, 286 p.
- Ingamells, C.O., 1974a, Control of geochemical error through sampling and subsampling diagrams: *Geochimica et Cosmochimica Acta*, v. 38, p. 1225-1237.
- Ingamells, C.O., 1974b, New approaches to geochemical analysis and sampling: *Talanta*, v. 24, p. 141-155.
- Ingamells, C.O., 1981, Evaluation of skewed exploration data—The nugget effect: *Geochimica et Cosmochimica Acta*, v. 45, p. 1209-1216.
- Ingamells, C.O., 1986, Sampling demonstration: Commonwealth Science Council Technical Publication Series, v. 213, p. 16-20.
- Ingamells, C.O., and Engels, J.C., 1976, Preparation, analysis, and sampling constants for a biotite, in *Accuracy in trace analysis: Sampling, sample handling, analysis*: National Bureau of Standards, Special Publication 422, p. 410-419.
- Ingamells, C.O., and Switzer, P., 1973, A proposed sampling constant for use in geochemical analysis: *Talanta*, v. 20, p. 547-568.
- Ingamells, C.O., Engels, J.C., and Switzer, P., 1972, Effect of laboratory sampling error in geochemistry and geochronology: *International Geological Congress, Report of the Geochemistry Section*, v. 24, no. 10, p. 405-415.
- Kleeman, A.W., 1967, Sampling error in the chemical analysis of rocks: *Journal of the Geological Society of Australia*, v. 14, p. 43-47.
- Lyman, G.J., 1998, The influence of segregation of particulates on sampling variance—The question of distributional heterogeneity: *International Journal of Mineral Processing*, v. 55, p. 95-112.
- Meyer, S.L., 1975, *Data analysis for scientists and engineers*: New York, Wiley, 513 p.
- Miller, R.E., 1991, The Duncan versus Visman conjecture as a predictor of trace element sampling variance in a heterogeneous body: The Sykesville Formation, Montgomery County, Maryland: *Eos, Transactions of the American Geophysical Union*, v. 72, no. 17, p. 308.
- Pitard, F.F., 1989a, Pierre Gy's sampling theory and sampling practice, 1: Boca Raton, CRC Press, 214 p.
- Pitard, F.F., 1989b, Pierre Gy's sampling theory and sampling practice, 2: Boca Raton, CRC Press, 247 p.
- Royle, A.G., 1985, A survey of gold sampling preparation methods: *Luma*, p. 48-54.
- Shaw, D.M., 1961, Manipulative errors in geochemistry: *Transactions of the Royal Society of Canada*, v. 60, p. 41-55.
- Sketchley, D.A., 1997, Case history guidelines for establishing sampling protocols and monitoring quality control: *Exploration and Mining Geology*, v. 6, no. 4, p. 384.
- Sketchley, D.A., 1998, Gold deposits: Establishing sampling protocols and monitoring quality control: *Exploration and Mining Geology*, v. 7, no. 1-2, p. 129-138.
- Smee, B.W., and Stanley, C.R., 2005, Sample preparation for nuggety samples: Dispelling some myths about sample size: *Explore, Newsletter of the Association of Exploration Geochemists*, no. 126, p. 21-27.
- Spiegel, M.R., 1975, *Probability and statistics—Schaum's outline series in mathematics*: New York, McGraw-Hill, 372 p.
- Stanley, C.R., 1990, Error propagation and regression on Pearce element ratio diagrams, in *Theory and application of Pearce element ratios to geochemical data analysis*, Russell, J.K., and Stanley, C.R., eds.: *Geological Association of Canada, Short Course Notes* 8, p. 179-216.
- Stanley, C.R., 1998, NUGGET: PC-software to calculate parameters for samples and elements affected by the nugget effect: *Exploration and Mining Journal, Canadian Institute of Mining and Metallurgy*, v. 7, no. 1-2, p. 139-147.
- Stanley, C.R., 2003, Estimating sampling errors for major and trace elements in geological materials using a propagation of variance approach: *Geochemistry: Exploration, Environment, Analysis*, v. 3, no. 2, p. 169-178.
- Stanley, C.R., 2006a, Numerical transformation of geochemical data, II: Stabilizing measurement error to facilitate data interpretation: *Geochemistry: Exploration Environment, Analysis*, v. 6, p. 79-96.
- Stanley, C.R., 2006b, On the special application of Thompson-Howarth error analysis to geochemical variables exhibiting a nugget effect: *Geochemistry: Exploration, Environment, Analysis*, v. 6, p. 357-368.
- Stanley, C.R., and Lawie, D., in press, Average relative error in geochemical determinations: Clarification, calculation, and a plea for consistency: *Exploration and Mining Geology*, in press.
- Stanley, C.R., and Smee, B.W., 2005, Reply to discussion by Dominique Francoise-Bongarcon regarding: Sample preparation for nuggety samples: Dispelling some myths about sample size: *Explore, Newsletter of the Association of Exploration Geochemists*, no. 127, p. 19-22.
- Stanley, C.R., and Smee, B.W., in press, The nature of sampling errors in exploration drilling programs for gold deposits: *Geochemistry: Exploration, Environment, Analysis*, in press.
- Visman, J., 1969, A general sampling theory: *Materials Research and Standards*, v. 9, p. 8-13.
- Visman, J., Duncan, A.J., and Lerner, M., 1971, Further discussion: A general sampling theory: *Materials Research and Standards*, v. 11, p. 32-37.
- Wickman, F.E., 1962, The amount of material necessary for a trace element analysis: *Arkiv för Mineralogi Och Geologi*, v. 10, p. 131-139.
- Wilson, A.D., 1964, The sampling of silicate rock powders for chemical analysis: *Analyst*, v. 89, p. 18-30.